Toward Instance-Optimal State Certification With Incoherent Measurements

Sitan Chen*
sitanc@berkeley.edu
UC Berkeley

Jerry Li
jerrl@microsoft.com
Microsoft Research

Ryan O'Donnell†
odonnell@cs.cmu.edu
Carnegie Mellon University

November 12, 2021

Abstract

We revisit the basic problem of quantum state certification: given copies of unknown mixed state $\rho \in \mathbb{C}^{d \times d}$ and the description of a mixed state σ, decide whether $\sigma = \rho$ or $\|\sigma - \rho\|_{\mathsf{tr}} \geq \varepsilon$. When σ is maximally mixed, this is *mixedness testing*, and it is known that $\Omega(d^{\Theta(1)}/\varepsilon^2)$ copies are necessary, where the exact exponent depends on the type of measurements the learner can make [OW15, BCL20], and in many of these settings there is a matching upper bound [OW15, BOW19, BCL20].

Can one avoid this $d^{\Theta(1)}$ dependence for certain kinds of mixed states σ, e.g. ones which are approximately low rank? More ambitiously, does there exist a simple functional $f : \mathbb{C}^{d \times d} \to \mathbb{R}_{>0}$ for which one can show that $\Theta(f(\sigma)/\varepsilon^2)$ copies are necessary and sufficient for state certification with respect to *any* σ? Such *instance-optimal* bounds are known in the context of classical distribution testing, e.g. [VV17].

Here we give the first bounds of this nature for the quantum setting, showing (up to log factors) that the copy complexity for state certification using nonadaptive incoherent measurements is essentially given by the copy complexity for mixedness testing times the fidelity between σ and the maximally mixed state. Surprisingly, our bound differs substantially from instance optimal bounds for the classical problem, demonstrating a qualitative difference between the two settings.

*This work was supported in part by NSF Award 2103300, NSF CAREER Award CCF-1453261, NSF Large CCF-1565235, and Ankur Moitra's ONR Young Investigator Award.

†Some of this work was done while the author was working at Microsoft Quantum. Supported by NSF grant FET-1909310 and ARO grant W911NF2110001. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

Contents

1 Introduction

We consider the problem of *quantum state certification*. We are given a description of a mixed state $\sigma \in \mathbb{C}^{d\times d}$ as well as N copies of a state $\rho \in \mathbb{C}^{d\times d}$. We are promised that either $\rho = \sigma$, or ρ is ε-far from σ in trace norm, and our goal is to distinguish between these two cases with high probability. From a practical perspective, the development of better methods for state certification is motivated by the need to efficiently verify the output of quantum devices. From a theoretical perspective, state certification is the natural quantum analogue of the well-studied classical problem of *identity testing*: given a description of a probability distribution p and samples from another distribution q, determine with high probability whether $q = p$ or $\|q - p\|_1 \geq \varepsilon$.

It is known that for general σ, $O(d/\varepsilon^2)$ copies of ρ suffice [BOW19]. Notably, this is smaller than the $\Theta(d^2/\varepsilon^2)$ copies needed to learn the state to ε-accuracy in trace norm. Prior work of [OW15] also demonstrated that when σ is the maximally mixed state, $\Omega(d/\varepsilon^2)$ copies are necessary [OW15]. While these results settle the copy complexity of this problem for worst-case choices of σ, they leave a number of interesting questions unanswered:

Using Incoherent Measurements. An important practical drawback of [BOW19] is that it makes a *coherent measurement* across the product state $\rho^{\otimes N}$. While such measurements are very powerful, they require the learner to keep all N copies of ρ in quantum memory without any of them decohering. In practice, creating such a large amount of quantum memory, even for medium sized d, has proven to be a difficult task, limiting the near-term viability of coherent measurements. In contrast, algorithms that make *incoherent measurements* only need to maintain one copy of ρ at a time. Additionally, whereas the measurement in [BOW19] takes $\mathrm{poly}(d, N)$ time to prepare, the protocol we present later in this paper can be implemented in $N \cdot \mathrm{poly}\log d$ time (see Remark 6.12). Understanding whether one achieve statistical guarantees similar to that of [BOW19] using only incoherent measurements is thus a crucial step towards reliable near-term quantum computation.

Recent work of [BCL20] studied this question in the special case where σ is the maximally mixed state– this special case of state certification is sometimes called *mixedness testing*. They showed that the practical viability of incoherent measurements unfortunately comes at a statistical cost: in this setting $\Omega(d^{4/3}/\varepsilon^2)$ copies are necessary, even if the incoherent measurements are chosen *adaptively* as a function of the previous measurement outcomes. When they are chosen *non-adaptively*, [BCL20] further showed that $\Theta(d^{3/2}/\varepsilon^2)$ copies are necessary *and sufficient*.

It is not too hard to modify their upper bound to show that for *general* σ, $O(d^{3/2}/\varepsilon^2)$ copies still suffice for state certification. This settles the copy complexity of state certification with non-adaptive, incoherent measurements for *worst-case* choices of σ.

Beyond Worst-Case σ. This raises another important question: for which σ can this $O(d^{3/2}/\varepsilon^2)$ upper bound be improved? This bound is certainly not tight for all σ: for instance, if σ is maximally mixed over a known subspace of dimension r, a simple argument demonstrates that $O(r^{3/2}/\varepsilon^2)$ copies suffice. A natural hypothesis might be that some relaxed notion of rank of σ dictates the true copy complexity of state certification with respect to σ.

This is inspired by a line of work in classical distribution testing on so-called *instance-optimal* bounds for identity testing [ADJ+11, ADJ+12, VV17, DK16, BCG19, JHW18]. The flagship result in this literature, due to [VV17], states that for any distribution p over d elements, the optimal sample complexity N of identity testing with respect to p is essentially characterized by the $\ell_{2/3}$-*quasinorm of* p. More formally, N satisfies:

$$\Omega(\varepsilon^{-1} \vee \varepsilon^{-2}\|p^{-\max}_{-\varepsilon/16}\|_{2/3}) \leq N \leq O(\varepsilon^{-1} \vee \varepsilon^{-2}\|p^{-\max}_{-\varepsilon}\|_{2/3})$$

for absolute constants $C_1, C_2 > 0$. Here $\|\cdot\|_{2/3}$ is the $\ell_{2/3}$-quasinorm, and $p^{-\max}_{-\delta}$ is the vector

given by zeroing out the largest entry as well as the bottom δ mass from the probability vector for p. Note that when p is uniform over d elements, this recovers the well-known sample complexity bound of $\Theta(\sqrt{d}/\varepsilon^2)$ for uniformity testing [Pan08].

Together, these two bounds give a striking and more or less tight characterization of the sample complexity landscape for identity testing: for *any instance* of the problem, we know the optimal sample complexity up to constant factors! This begs the natural question:

Can we get a similarly tight characterization for the copy complexity of state certification with incoherent measurements?

1.1 Our Results

In this work, we answer this in the affirmative by presenting an instance-optimal characterization of the copy complexity of state certification with non-adaptive incoherent measurements. Surprisingly, our results demonstrate that the behavior of quantum state certification is qualitatively quite different from that of classical identity testing. More formally, our main result is the following:

Theorem 1.1 (Informal, see Theorems 5.1 and 6.1)**.** *Given any mixed state $\sigma \in \mathbb{C}^{d\times d}$, there are mixed states $\overline{\sigma}$ and $\underline{\sigma}$ respectively given by projecting away some eigenvectors with eigenvalues summing to at most $\Theta(\varepsilon^2)$ and $\Theta(\varepsilon)$ and normalizing, such that the following holds.*

Let $\overline{d}_{\mathsf{eff}}$ (resp. $\underline{d}_{\mathsf{eff}}$) be the rank of $\overline{\sigma}$ (resp. $\underline{\sigma}$). The optimal copy complexity N of state certification with respect to σ to trace distance ε using non-adaptive, incoherent measurements satisfies[1]

$$\widetilde{\Omega}\left(\frac{d\cdot \underline{d}_{\mathsf{eff}}^{1/2}}{\varepsilon^2}\cdot F(\underline{\sigma},\rho_{\mathsf{mm}})\right)\le N\le \widetilde{O}\left(\frac{d\cdot \overline{d}_{\mathsf{eff}}^{1/2}}{\varepsilon^2}\cdot F(\overline{\sigma},\rho_{\mathsf{mm}})\right),$$

where ρ_{mm} is the maximally mixed state $\frac{1}{d}\mathbb{1}$ and F denotes the fidelity between two quantum states.

Note that when σ is maximally mixed and $0<\varepsilon<1$ is bounded away from 1, then $\overline{\sigma}$ and $\underline{\sigma}$ are projectors to subspaces of dimension $\Omega(d)$, so $\underline{d}_{\mathsf{eff}},\overline{d}_{\mathsf{eff}}=\Theta(d)$ and $F(\underline{\sigma},\rho_{\mathsf{mm}})=\Theta(1)$, recovering[2] the $\Theta(d^{3/2}/\varepsilon^2)$ bound of [BCL20] for mixedness testing with non-adaptive, incoherent measurements.

Qualitatively, our result says that unless σ puts $1-\mathrm{poly}(\varepsilon)$ mass on $o(d)$ dimensions, the copy complexity of state certification is equal to the worst-case copy complexity of state certification, times the fidelity between σ and the maximally mixed state. Surprisingly, unlike in the classical case, our bound demonstrates that there is no clean dimension-independent functional which controls the complexity of quantum state certification. Rather, there is some inherent "curse of dimensionality" for this problem. Also note that in the quantum case, unlike in the classical case, we do not remove the largest element from the spectrum of σ.

Example 1.2. *To elaborate on this curse of dimensionality, consider the following example. Let $\sigma\in\mathbb{C}^{(d+1)\times(d+1)}$ be the mixed state given by $\sigma=\mathrm{diag}(1-1/d^2,1/d^3,\dots,1/d^3)$. The classical analogue of certifying this state is identity testing to the distribution p over $d+1$ elements which has one element with probability $1-1/d^2$, and d elements with probability $1/d^3$.*

For the classical case, the bound from [VV17] demonstrates that the sample complexity of identity testing to p is $\Theta\left(\frac{1}{d^{3/2}\varepsilon^2}\right)$ for sufficiently small ε. In particular, in this regime the sample complexity actually is decreasing *in d. This phenomena is not too surprising—this distribution is very close*

[1] Throughout, we use $\widetilde{\Omega}(\cdot)$ and $\widetilde{O}(\cdot)$ solely to suppress factors of $\log(d/\varepsilon)$.

[2] As our techniques are a strict generalization of those of [BCL20], in this special case where σ is the maximally mixed state, our analysis does not actually lose log factors.

to being a point distribution, and the only "interesting" part of it, namely, the tail, only has total mass $1/d$, which vanishes as we increase d.

In contrast, Theorem 1.1 shows that the copy complexity of the quantum version of this problem using incoherent measurements is $\widetilde{\Theta}(d^{1/2}/\varepsilon^2)$. Notably, this is increasing *in d! At a high level (see Section 2 for further discussion), it is because the unknown state ρ may share the same diagonal entries with σ but may not commute with it, so the "interesting" behavior need not be constrained to the subspace given by the small eigenvalues of σ. In particular, ρ might be far from σ only because ρ contains nontrivial mass in its off-diagonal entries. This allows us many more degrees of freedom in constructing the lower bound instance, resulting in a much stronger bound.*

It turns out this curse of dimensionality persists even for *adaptive*, incoherent measurements. Formally, we show the following lower bound which is qualitatively similar to that of Theorem 1.1:

Theorem 1.3 (Informal, see Theorem A.1)**.** *In the notation of Theorem 1.1,*

$$N \geq \widetilde{\Omega}\left(\frac{d \cdot \underline{d}_{\mathsf{eff}}^{1/3}}{\varepsilon^2} \cdot F(\underline{\sigma}, \rho_{\mathsf{mm}})\right) \tag{1}$$

copies are needed for state certification w.r.t. σ to error ε using adaptive, incoherent measurements.

When $\sigma = \rho_{\mathsf{mm}}$, we recover the best known adaptive lower bound for mixedness testing [BCL20]. Furthermore, since non-adaptive measurements are a subset of adaptive ones, the upper bound in Theorem 1.1 also provides a per-instance upper bound for this problem which matches (1) up to the factor of $d^{1/2}$ versus $d^{1/3}$. Obtaining tight bounds in this setting is an interesting open question; however, we note that this is not known even for mixedness testing.

1.2 Related Work

A full survey of the literature on quantum (and classical) testing is beyond the scope of this paper; we only discuss the most relevant works below. We also note there is a vast literature on related quantum learning problems such as state tomography, see e.g. [KRT17, GLF+10, FGLE12, Vor13, HHJ+17, OW16, OW17] and references therein.

Quantum Property Testing. The problem of quantum state certification lies within the broader field of quantum state property testing. See [MdW16] for a more complete survey. Within this field, there are two regimes studied. In the *asymptotic* regime, the goal is to precisely characterize the rate at which the error converges as $n \to \infty$, and d and ε are fixed. Here quantum state certification is more commonly known as *quantum state discrimination* [Che00, BC09, ANSV08]. For a more complete survey of work on this problem, see [BK15]. However, this line of work does not attempt to characterize the statistical dependence on the dimension.

In contrast, we consider the *non-asymptotic* regime, where the goal is to characterize the rate of convergence for quantum state certification as a function of d and ε. As discussed above, recent work of [OW15, BOW19] has demonstrated that $\Theta(d/\varepsilon^2)$ copies are necessary and sufficient for quantum state certification over the worst choice of σ, when the measurements are allowed to be arbitrary. However, the representation theoretic tools used within seem to be quite brittle and do not easily extend to give instance-optimal rates. Understanding the instance-optimal rate for quantum state certification using arbitrary measurements is a very interesting open question.

Incoherent Measurements. A number of recent papers on quantum learning and testing have also considered the power of incoherent measurements, and more generally, other types of restricted measurements for quantum property testing tasks apart from state certification. Following the

aforementioned [BCL20], subsequent work of [ACQ21] defined a more general notion of quantum algorithmic measurement which includes incoherent measurements and proved some incomparable lower bounds for other problems such as purity testing and channel discrimination under this model. The recent work of [HKP21] also showed a separation for shadow tomography with Pauli observables using incoherent versus 2-entangled measurements. Lastly, another very recent work [CCHL21] refined these two works by showing nearly optimal separations for shadow tomography, purity testing, and channel discrimination using incoherent versus coherent measurements.

We also note that a number of papers in quantum tomography have considered the power of incoherent measurements, see e.g. [KRT17, GLF+10, FL11, Vor13, HHJ+17]. Another line of work considers the complexity of testing using only Pauli measurements [FL11, FGLE12, dSLCP11, AGKE15]. However, because of the restrictive setting, these latter bounds are typically weaker, and these papers also do not obtain instance-optimal bounds for this setting.

Classical Distribution Testing. State certification is the quantum version of the well-studied classical problem of distribution identity testing. A complete survey of this field is also beyond the scope of this paper. See [Can20, Gol17] and references within for a more detailed discussion. Of particular interest to us is the line of work on instance-optimal testing, the direct classical analog of the problem we consider in this paper. The works of [ADJ+11, ADJ+12] consider sample complexity bounds which improve upon the worst case sample complexity for different choices of probability distributions. The setting that we consider is most directly inspired by the aforementioned work of [VV17]. Subsequent work has re-proven and/or derived new instance-optimal bounds for identity testing and other problems as well, see e.g. [DK16, BCG19, JHW18].

2 Overview of Techniques

As with many other property testing lower bounds, ours is based on showing hardness for distinguishing between a simple "null hypothesis" and a "mixture of alternatives," i.e. whether the unknown state ρ that we get copies of is equal to σ or was randomly sampled at the outset from some distribution over states ε-far from σ. Throughout, we will assume that σ is a diagonal matrix. This is without loss of generality since we are given a description of σ and can change basis.

When $\sigma = \frac{1}{d}\mathbb{1}$, the standard choice for the mixture (and the one that leads to optimal lower bounds in this case) is the distribution over mixed states of the form $\frac{1}{d}\left(\mathbb{1} + \mathbf{U}^\dagger \mathrm{diag}(\varepsilon, \ldots, -\varepsilon, \ldots)\mathbf{U}\right)$ where $\mathbf{U}$ is sampled from the Haar measure over $d \times d$ unitary matrices, and previous works have shown lower bounds for mixedness testing with entangled measurements [OW15] and incoherent measurements [BCL20] by analyzing this particular distinguishing task. Indeed, our proof builds upon the general framework introduced in the latter work (see Section 4 for an exposition of the main ingredients from [BCL20]) but differs in crucial ways.

To get a sense for what the right distinguishing task(s) to consider for general σ are, it is instructive to see first how to prove instance-optimal bounds for classical distribution testing.

2.1 Instance-Optimal Lower Bounds for Identity Testing

Here we sketch how to prove the lower bound of [VV17] for identity testing (up to log factors). Recall this is the setting where one gets access to independent samples from an unknown distribution p over d elements and would like to test whether $p = q$ or $\|p - q\|_1 > \varepsilon$ for a known distribution q.

When q is the uniform distribution over d elements, a classical result of [Pan08] demonstrates that the fundamental bottleneck is distinguishing whether the samples come from p, or if the samples come from a version of q where each entry from its vector of probabilities has been perturbed by

$\pm\varepsilon/d$. In this setting, the mixture of alternatives consists of all distributions q^ζ that could have been obtained in this fashion, where the index ζ indicates the sign pattern of the perturbation chosen.

The main conceptual challenge to extending this lower bound strategy to more general q is that the entries of the probability vector for q may take values across many different scales, and whatever lower bound instance one designs must be sensitive to these scales.

One approach to account for these different scales is to "bucket" the probability vector for q, where each given bucket contains all entries within a fixed multiplicative factor of one another. It turns out that Paninski's analysis works even if q is not exactly uniform as long as its probabilities are within a multiplicative factor of each other. For this reason, within each bucket we could simply apply Paninski's construction and randomly perturb the probabilities by a carefully chosen multiple of $\pm\varepsilon/d$. Combining these constructions across buckets after appropriately scaling them thus gives a natural mixture of alternatives $\{q^\zeta\}$ to distinguish from the true distribution q, where again, ζ denotes the sign pattern of the perturbations chosen.

The main technical challenge then is to upper bound $d_{\mathrm{TV}}(q^{\otimes N}, \mathbb{E}_\zeta[(q^\zeta)^{\otimes N}])$, that is, the total variation distance between the distribution over N i.i.d. draws from q and the distribution over N i.i.d. draws from q^ζ where ζ was sampled uniformly at random from the set of all possible sign patterns corresponding to perturbations of q.

A common analytical trick for carrying out this bound— and the approach that [VV17] take— is to first Poissonize, that is, take N to be a Poisson random variable. Unfortunately, Poissonization does not seem to have any straightforward analogue in the quantum setting, where the choice of measurement can vary across copies, so we eschew this technique in favor of an alternative approach that we sketch next.

Ingster-Suslina Method and Moment Bounds. Apart from Poissonization, another way to bound $d_{\mathrm{TV}}(q^{\otimes N}, \mathbb{E}_\zeta[(q^\zeta)^{\otimes N}])$ is to pass to chi-squared divergence and invoke the Ingster-Suslina method (see e.g. Section 3.3 of [IS12], or Lemma 22.1 and its application in Section 24.3 in [Wu17]). At a high level, this approach amounts to bounding higher-order moments of the *pairwise correlation*

$$\phi^{\zeta,\zeta'} \triangleq \mathop{\mathbb{E}}_i\big[(\Delta_\zeta(i)-1)(\Delta_{\zeta'}(i)-1)\big]$$

as a random variable in ζ, ζ'. Here, the expectation is over sample $i \in [d]$ drawn from q, and

$$\Delta_\zeta(i) = q_i^\zeta / q_i$$

is the *likelihood ratio* between the probability of drawing i when $p = q^\zeta$ versus the probability of drawing i when $p = q$. Concretely, if one can show that

$$\mathop{\mathbb{E}}_{\zeta,\zeta'}\left[\left(1+\phi^{\zeta,\zeta'}\right)^t\right] = 1 + o(1)$$

for some t, this would imply a sample complexity lower bound of t for testing identity to q.

It turns out to be possible to give sufficiently good upper bounds on the moments of $\phi^{\zeta,\zeta'}$ (after some appropriate preprocessing on q as done in [VV17]) that one can recover the same bound as [VV17] up to poly-logarithmic factors in d/ε. It is this approach that we will generalize to the quantum setting.

2.2 Passing to the Quantum Setting

We now describe how to extend some of these ideas to quantum state certification.

Scale-Sensitive Rotations. Recall from the discussion at the beginning of this section that in the case where $\sigma = \rho_{\mathsf{mm}}$, the right "mixture of alternatives" to consider is to perturb every eigenvalue of ρ_{mm} and then randomly rotate by a Haar-random unitary over $\mathbb{C}^d$; this is sometimes called the *quantum Paninski* instance [OW15] for its resemblance to Paninski's construction in the classical setting.

For general σ, we could try the same thing, but motivated by the classical setting, we would tune how much we perturb each eigenvalue based on its magnitude. Unfortunately, if we then simply rotate the resulting perturbed state by a Haar-random unitary over $\mathbb{C}^d$, it turns out that we can't hope to prove a sufficiently strong lower bound.

To see this, let's consider the following extreme example. Imagine that σ is nearly a pure state. A random global rotation of a perturbation of σ, no matter how cleverly we picked the perturbation, is close to a Haar-random pure state. So its trace inner product with σ will be on the order of $\Theta(1/d)$ with high probability, whereas the trace inner product of σ with itself is on the order of $\Theta(1)$. In particular, just by measuring the observable given by σ, we can easily distinguish whether $\rho = \sigma$ or ρ comes from this particular mixture of alternatives using $O(1)$ measurements.

The point is that in the quantum setting, we need to be sensitive to the different scales of σ's eigenvalues not only in picking the perturbations to the eigenvalues of σ, but *also in picking the ensemble of rotations*!

An Attempt: Generalized Quantum Paninski. We now outline an attempt at generalizing the quantum Paninski construction in a way that is sufficiently sensitive to the different scales for the eigenvalues of σ. Motivated by the classical construction described above, we can group the eigenvalues of σ into *buckets*, where a given bucket contains all eigenvalues within a fixed multiplicative factor of each other, and consider a mixture of alternatives defined as follows. First, given any $m \in \mathbb{N}$, define the matrix:

$$\mathbf{Z}_m \triangleq \begin{cases} \mathrm{diag}(1, \ldots, -1, \ldots) & \text{m even} \\ \mathrm{diag}(0, 1, \ldots, -1, \ldots) & \text{m odd,} \end{cases}$$

where $\mathbf{Z}_m$ consists of $\lfloor m/2 \rfloor$ 1's and $\lfloor m/2 \rfloor$ -1's. The mixture of alternatives is given by the distribution over mixed states of the form $\sigma + \mathbf{U}^\dagger \mathcal{E} \mathbf{U}$, where now $\mathbf{U}$ is a *block-diagonal unitary matrix whose blocks are Haar-random* and whose block structure corresponds to the buckets, and $\mathcal{E}$ is a direct sum of scalings of $\mathbf{Z}_m$, where the different m's and scalings correspond to the sizes and relative magnitudes of the buckets.

For instance, if $\sigma = \left(\frac{1}{2\sqrt{d}} \mathbb{1}_{\sqrt{d}}\right) \oplus \left(\frac{1}{2(d-\sqrt{d})} \mathbb{1}_{d-\sqrt{d}}\right)$, we can take $\mathbf{U}$ to be distributed as $\mathbf{U}_1 \oplus \mathbf{U}_2$, where $\mathbf{U}_1 \in U(\sqrt{d})$ and $\mathbf{U}_2 \in U(d - \sqrt{d})$ are Haar-random, and $\mathcal{E} = \left(\frac{\varepsilon_1}{2\sqrt{d}} \mathbf{Z}_{\sqrt{d}}\right) \oplus \left(\frac{\varepsilon_2}{2(d-\sqrt{d})} \mathbf{Z}_{d-\sqrt{d}}\right)$ for appropriately chosen $\varepsilon_1, \varepsilon_2$ summing to 2.

Our analysis for this instance follows the Ingster-Suslina method in the nonadaptive case and the general framework of [BCL20] in the adaptive case (see Section 4 for an exposition of these two frameworks), and the central object for both proofs is the pairwise correlation

$$\phi^{\mathbf{U},\mathbf{V}} \triangleq \mathop{\mathbb{E}}_{z}[(\Delta_{\mathbf{U}}(z) - 1)(\Delta_{\mathbf{V}}(z) - 1)].$$

Analogously to the classical setup described above, here the expectation is over outcomes z if one makes some quantum measurement on a single copy of the state $\rho = \sigma$, and $\Delta_{\mathbf{U}}(z)$ is the likelihood ratio between the probability of observing outcome z when $\rho = \sigma + \mathbf{U}^\dagger \mathcal{E} \mathbf{U}$ versus the probability

of observing the same outcome when $\rho = \sigma$ under a particular POVM (see Section 4 for formal definitions). And as in the classical setup, it turns out that we need to show that

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\left(1+\phi^{\mathbf{U},\mathbf{V}}\right)^t\right] = 1 + o(1)$$

for sufficiently large t, so the primary challenge is to control the moments of $\phi^{\mathbf{U},\mathbf{V}}$ (regarded as a random variable in $\mathbf{U}, \mathbf{V}$), or equivalently to show that it concentrates sufficiently around its mean.

If $\mathbf{U}, \mathbf{V}$ were Haar-random unitary matrices, one could do this by invoking standard concentration of measure for Haar-random unitary matrices [AGZ10, MM13]. Indeed, this is the approach of [BCL20], but for general σ we need to control the tails of $\phi^{\mathbf{U},\mathbf{V}}$ when $\mathbf{U}, \mathbf{V}$ have the above-mentioned block structure, for which off-the-shelf tail bounds will not suffice. Instead, we argue that because we can assume without loss of generality that the optimal measurements to use to distinguish $\rho = \sigma$ from $\rho = \sigma + \mathbf{U}^\dagger \mathcal{E} \mathbf{U}$ must respect the block structure, $\phi^{\mathbf{U},\mathbf{V}}$ is a weighted sum of pairwise correlations $\phi_j^{\mathbf{U},\mathbf{V}}$ for many independent sub-problems, one for each "bucket" j (see (9)). These are independent random variables, each parametrized by an independent Haar-random unitary matrix in a lower-dimensional space, so we can show a tail bound for $\phi^{\mathbf{U},\mathbf{V}}$ by combining the tail bounds for $\{\phi_j^{\mathbf{U},\mathbf{V}}\}$ (see Section 5.2.1).

In Section 5.2.2, we show how to optimally tune the entries of $\mathcal{E}$. Here however, we finally arrive at the surprising juncture where instance-optimal state certification deviates significantly from its classical analogue:

This generalized quantum Paninski construction does not always yield the right lower bound!

It turns out that even with the optimal tuning of $\mathcal{E}$, the approach outlined thus far only achieves a copy complexity lower bound of roughly $\widetilde{\Omega}(\|\sigma'\|_{2/5}/\varepsilon^2)$ (see Lemma 5.5), where σ' is obtained from σ by projecting out its largest eigenvalue and some small eigenvalues.

While this recovers the lower bound of [BCL20] when $\sigma = \rho_{\mathsf{mm}}$, in other situations one can readily see that $\widetilde{\Omega}(\|\sigma'\|_{2/5}/\varepsilon^2)$ can be much worse than the lower bound in Theorem 1.1. Consider σ given by Example 1.2. For that choice of $\sigma = \mathrm{diag}(1 - 1/d^2, 1/d^3, \ldots, 1/d^3)$, $\|\sigma'\|_{2/5} = 1/\sqrt{d}$, so we only get a lower bound of $\Omega\left(\frac{1}{d^{1/2}\varepsilon^2}\right)$. In contrast, as discussed in Example 1.2, the right copy complexity for this problem turns out to be $\widetilde{\Theta}(\sqrt{d}/\varepsilon^2)$. We now describe a second lower bound instance that, combined with the generalized quantum Paninski construction, yields an instance near-optimal lower bound.

Missing Ingredient: Perturbing the Off-Diagonals. For simplicity, consider a mixed state σ with exactly two buckets, e.g. $\sigma = (\lambda_1 \mathbb{1}_{d_1}) \oplus (\lambda_2 \mathbb{1}_{d_2})$ where $d_1 \geq d_2$. In this case, one can regard the generalized Paninski instance as a family of perturbations of the two principal submatrices indexed by the coordinates $\{1, \ldots d_1\}$ in bucket 1 and the coordinates $\{d_1 + 1, \ldots, d\}$ in bucket 2 respectively. But one could also perturb σ along the *off-diagonal blocks*, rather than on the principal blocks, by considering matrices of the form

$$\sigma + \begin{pmatrix} \mathbf{0}_{d_1} & (\varepsilon/2d_2) \cdot \mathbf{W} \\ (\varepsilon/2d_2) \cdot \mathbf{W}^\dagger & \mathbf{0}_{d_2} \end{pmatrix} \tag{2}$$

parametrized by Haar-random $\mathbf{W} \in \mathbb{C}^{d_1 \times d_2}$ consisting of orthonormal columns. One can show that as long as $\varepsilon \leq d_{j_1} \cdot \sqrt{\lambda_1 \lambda_2}$, then (2) is a valid density matrix (Lemma 5.18) and is ε-far in trace distance from σ. In this regime, we show a lower bound of $\Omega(d_1\sqrt{d_2}/\varepsilon^2)$ for distinguishing whether $\rho = \sigma$ or whether ρ is given by a matrix (2) where $\mathbf{W}$ is sampled Haar-randomly at the outset.

For general σ, by carefully choosing which pair of buckets to apply this construction to, we obtain the lower bound of Theorem 1.1 for very small ε. For larger ε we show that if the lower bound from the generalized Paninski instance were inferior to that of Theorem 1.1, then this would contradict the assumption that ε is large (see Section 5.5). Altogether, this completes the proof of the claimed lower bound in Theorem 1.1, modulo one last corner case that we now discuss.

Handling the Largest Eigenvalue. Indeed, there is one more feature of Theorems 1.1 and 1.3 which is unique to the quantum setting. In the classical setting, the instance-optimal sample complexity of testing identity to a given distribution p is essentially given by $\frac{1}{\varepsilon} \vee \frac{\|p'\|_{2/3}}{\varepsilon^2}$, where p' is derived from p by zeroing out not just the bottom $O(\varepsilon)$ mass from p but also the *largest* entry of p. To see why the latter, as well as the additional $\frac{1}{\varepsilon}$ term, is necessary, consider a discrete distribution p which places $1 - \varepsilon/100$ mass on some distinguished element of the domain, call it x^*. The $\frac{1}{\varepsilon} \vee \frac{\|p'\|_{2/3}}{\varepsilon^2}$ lower bound would yield $\Omega(1/\varepsilon)$ sample complexity, and an algorithm matching this bound would simply be to estimate the mass the unknown distribution places on x^*. The reason is that because p places total mass $\varepsilon/100$ on elements distinct from x^*, any distribution ε-far from p in ℓ_1-distance must place at most $1 - \varepsilon$ mass on x^*, which can be detected in $O(1/\varepsilon)$ samples.

In stark contrast, in the quantum setting if σ had an eigenvalue of $1 - \varepsilon/100$, then the copy complexity of state certification with respect to σ scales with $1/\varepsilon^2$. The reason is that there is "room in the off-diagonal entries" for a state ρ to be ε-far from σ. Indeed, we can formalize this by considering a lower bound instance similar to (2). In fact it is even simpler, because for mixed states whose largest eigenvalue is particularly large, it suffices to randomly perturb a single pair of off-diagonal entries! To analyze the resulting distinguishing task, we eschew the framework of [BCL20] and directly bound the likelihood ratio between observing any given sequence of measurement outcomes under the alternative hypothesis versus under the null hypothesis (see Section 5.4 and Lemma 5.24 in particular).

Adaptive Lower Bounds. As we discussed following Theorem 1.3, the ideas above can also be implemented in the setting where one can choose incoherent measurements *adaptively* (see Theorem 1.3). The reason the lower bound we obtain is not instance-optimal is the same technical reason that [BCL20] was not able to obtain an optimal lower bound in the special case of mixedness testing, namely that there is some lossy balancing step to handle a certain low-probability event (see the proof of Theorem 4.8 in Appendix B.2).

2.3 Upper Bound

As in our lower bound proof, we will partition the spectrum of σ into buckets. We will also place all especially small eigenvalues of σ in a single bucket of their own– this latter bucket will contain the smallest eigenvalues of σ that together sum to $O(\varepsilon^2)$. For the purposes of discussion in this section, we will call this the "negligible bucket" and we will call all others "non-negligible buckets."

For starters, in Section 6.1 we give a simple algorithm (BasicCertify, see Algorithm 1) for state certification which is already optimal up to constant factors when the eigenvalues of σ all fall within the same bucket. Similar to the mixedness tester in [BCL20], this algorithm is based on measuring our copies of unknown state ρ in a Haar-random basis and running a classical identity tester [DK16]. As the analysis is very similar to that of [BCL20], we defer the details to Section 6.1.

Now consider a general mixed state σ given by an arbitrary diagonal density matrix. Suppose its diagonal entries fall into m buckets in total; by virtue of the bucketing scheme, m is guaranteed to be at most logarithmic in d/ε (see Fact 5.3). At a high level, if the state ρ that we get copies

of is ε-far in trace distance from σ, then by an averaging argument, there should be some pair of buckets such that the corresponding block submatrix of σ is somewhat far from ρ in trace distance. Indeed, one of four things could happen (see Figure 1):

(A) There may be a non-negligible bucket for which the corresponding principal submatrix of σ is $\Omega(\varepsilon/m^2)$-far from that of ρ, in which case we can detect that ρ is far from σ simply by running BasicCertify restricted to that bucket (see Lemma 6.10).

(B) There may be two non-negligible buckets for which the corresponding pair of off-diagonal blocks in σ are $\Omega(\varepsilon/m^2)$-far from the corresponding submatrix in ρ, in which case we can detect that ρ is far from σ by running BasicCertify restricted to these two buckets (see Lemma 6.11).

(C) For the negligible bucket, the corresponding principal submatrix of σ is $\Omega(\varepsilon^2)$-far from that of ρ, in which case we can measure the observable given by the projector to that submatrix. In this case, $O(1/\varepsilon^2)$ copies suffice (see Lemma 6.8).

(D) None of the above three cases hold, and ρ and σ differ primarily in the off-diagonal block with rows indexed by the negligible bucket and columns indexed by all non-negligible buckets. But by basic linear algebra (Lemma 3.15) and the fact that the eigenvalues in the negligible bucket sum to ε^2, this would contradict the fact that we are not in case (C) (see Lemma 6.9)!

We remark that the idea of reducing from state certification to mixedness testing by performing a case analysis on buckets of the spectrum is reminiscent of the instance near-optimal algorithm of [DK16] for classical identity testing. That said, as is clear in the above proof sketch, the off-diagonal entries of ρ pose a number of technical hurdles not present in the classical setting, just as they did in the proof of the lower bound.

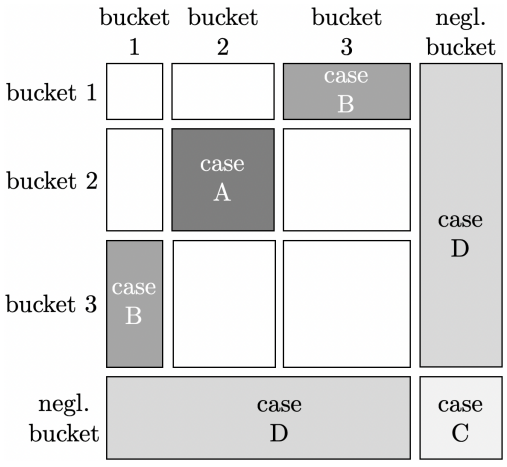

Figure 1: Partition of σ into blocks corresponding to buckets, relevant submatrix for each case highlighted in gray.

Why Do the Upper and Lower Bounds "Line Up"? The casework above gives a good sense for why our upper and lower bounds happen to "line up" up to log factors. Ignoring the negligible bucket for the time being, recall that the averaging argument in our upper bound proof essentially implies that any ρ which is far from σ must be relatively far either 1) within a principle submatrix corresponding to a single bucket, or 2) within an off-diagonal submatrix corresponding to a pair of buckets.

This upper bound strategy complements our lower bound constructions nicely. Indeed, if we ignore the contribution from all other entries apart from the submatrix in question, then we can ask: what mixture of alternatives is hardest to tell apart from σ if the alternatives all differ from σ only in that submatrix? Depending on whether that submatrix is principle or off-diagonal, our generalized quantum Paninski and off-diagonal lower bound constructions provide essentially the optimal answer to this question.

Why Truncation? The reader might be wondering why we need to truncate some of the eigenvalues of σ in our bounds in Theorems 1.1 and 1.3. For instance, how are we able to prove an

upper bound which only depends on σ after we have thrown out $O(\varepsilon^2)$ of its eigenmass, rather than on σ itself? As the above description of our algorithm makes clear, the reason is that the copy complexity of state certification with respect to σ is really dominated by cases (A) and (B), and in these cases the complexity of running BasicCertify only depends on the non-negligible buckets, i.e. the buckets containing the eigenvalues corresponding to the truncation $\overline{\sigma}$.

That said, there is a gap between the amount of mass we need to truncate in the definition of $\overline{\sigma}$ in the upper bound versus $\underline{\sigma}$ in the lower bound ($\Theta(\varepsilon^2)$ versus $\Theta(\varepsilon)$) in our theorems. The latter level of truncation appears to be an artifact of our techniques, and we conjecture that the lower bound can be upgraded to hold even if $\underline{\sigma}$ is defined by removing only $\Theta(\varepsilon^2)$ mass from σ.

Why Fidelity? Finally, we give some intuition for why fidelity with respect to the maximally mixed state arises in our copy complexity bounds. To do so, we will go into slightly more detail about the analysis of the algorithm we sketched above, focusing on cases (A) and (B).

First consider case (A). Suppose for simplicity that ρ was identical to σ except in the principal submatrix corresponding to the diagonal entries of σ in the interval $[2^{-j-1}, 2^{-j}]$. Denote the number of rows/columns of this submatrix by d_j. As we alluded to above, it turns out that mixedness testing to error ε' for d_j-dimensional mixed states whose eigenvalues are all in the same bucket has copy complexity $\Theta(d_j^{3/2}/\varepsilon'^2)$. On the other hand, because the trace of this submatrix is $\Theta(d_j 2^{-j})$, we would need to make $\Theta(2^j/d_j)$ measurements of ρ in expectation to simulate one measurement of the conditional state given by ρ restricted to this submatrix. But for the same reason, the trace distance ε' between the *normalized* states given by this principal submatrix of ρ and σ is also $2^j/d^j$ times bigger than ε/m^2. As m is logarithmic in d/ε, this means that $\widetilde{O}(d_j^{5/2}2^{-j}/\varepsilon^2)$ copies suffice to detect that ρ differs noticeably from σ in this submatrix (see Lemma 6.10).

Now consider case (B). Suppose for simplicity that ρ was identical to σ except in the $d_j \times d_{j'}$ and $d_{j'} \times d_j$ off-diagonal blocks corresponding to two buckets of eigenvalues, namely those in $[2^{-j-1}, 2^{-j}]$ and those in $[2^{-j'-1}, 2^{-j'}]$ (here $d_j, d_{j'}$ denote the sizes of these buckets). Also suppose without loss of generality that $d_j \geq d_{j'}$. It turns out that if we ran BasicCertify restricted to the $(d_j + d_{j'}) \times (d_j + d_{j'})$ principal submatrix of ρ containing these off-diagonal blocks, then by a more involved version of the reasoning in the previous paragraph (see Lemma 6.11), we can show that $\widetilde{O}(\sqrt{d_j} d_{j'}^2 2^{-j'}/\varepsilon^2)$ measurements of ρ suffice to detect that ρ differs noticeably from σ.

Putting everything together, we conclude that our algorithm needs to make, up to log factors,

$$\max_{j,j':d_j \geq d_{j'}} \sqrt{d_j} d_{j'}^2 2^{-j'}/\varepsilon^2 = \left(\max_j \sqrt{d_j}\right) \cdot \left(\max_{j'} d_{j'}^2 2^{-j'}\right)/\varepsilon^2 \tag{3}$$

measurements, where j, j' range over non-negligible buckets, $j = j'$ corresponds to case (A), and $j \neq j'$ corresponds to case (B). As there are logarithmically many nonempty non-negligible buckets of eigenvalues of σ, it is elementary to check that (3) is, up to log factors, equal to $\overline{d}_{\mathsf{eff}}^{1/2} \cdot \|\overline{\sigma}\|_{1/2}$ (see Fact 3.18), where $\|\cdot\|_{1/2}$ denotes the Schatten 1/2-quasinorm. Finally we can see where the fidelity term comes from: for any density matrix $\overline{\sigma} \in \mathbb{C}^{d \times d}$,

$$d \cdot F(\overline{\sigma}, \mathbb{1}/d) = d \cdot \frac{1}{d} \operatorname{Tr}(\overline{\sigma}^{1/2})^2 = \|\overline{\sigma}\|_{1/2},$$

Thus far we have only provided justification for why fidelity emerges in the upper bound. But as we mentioned in our discussion for why the upper and lower bounds happen to "line up," our generalized quantum Paninski and off-diagonal lower bound constructions closely parallel case (A) and case (B) in the upper bound analysis. Naturally, we end up seeing the same kinds of terms, e.g. $\|\cdot\|_{2/5}$ and $\|\cdot\|_{1/2}$, emerge in the proof of the lower bound for essentially the same reasons.

Roadmap In Section 3, we review basic notions in quantum property testing and present various technical tools we will use in our proofs. In Section 4 we describe the general framework introduced in [BCL20] for proving lower bounds with incoherent measurements. In Section 5, we prove the lower bound in Theorem 1.1, and in Section 6 we prove the upper bound. In Appendix A we prove Theorem 1.3. In Appendix B we collect some deferred proofs from the main body.

3 Technical Preliminaries

Notation Let S_ℓ denote the symmetric group on ℓ elements. Given $\pi \in S_\ell$, let $\kappa(\pi)$ denote the number of cycles in π. Recall from the introduction that we let $\rho_{\mathsf{mm}} \triangleq \frac{1}{d}\mathbb{1}$ denote the maximally mixed state. Given a matrix M and $p > 0$, let $\|M\|_p$ denote the Schatten-p (quasi)norm. Let $\widehat{M} \triangleq M/\operatorname{Tr}(M)$. Let $U(d)$ denote the unitary group of $d \times d$ matrices.

3.1 Quantum Property Testing

We will work with the following standard notions, using notation and terminology borrowed from [BCL20].

Definition 3.1. *A positive operator-valued measurement (POVM) $\mathcal{M}$ consists of a collection of psd matrices $M_1, ..., M_m$ for which $\sum M_i = \mathbb{1}$. We will refer to the set of measurement outcomes $[m]$ as $\Omega(\mathcal{M})$. Given mixed state ρ, the* distribution over outcomes from measuring ρ with $\mathcal{M}$ *is the distribution over $\Omega(\mathcal{M})$ which places mass $\langle M_i, \rho\rangle$ on outcome i.*

As demonstrated in [BCL20], the techniques in that work and in the present paper generalize easily to POVMs for which $\Omega(\mathcal{M})$ is infinite, so for simplicity we will simply consider the finite case in this work.

Definition 3.2. *Let $N \in \mathbb{N}$. A* POVM schedule *$\mathcal{S}$ is a collection of POVMs $\{\mathcal{M}^{x_{<t}}\}_{t\in[N], x_{<t}\in\mathcal{T}_t}$, where each $\mathcal{M}^{x_{<t}}$ is over $\mathbb{C}^d$, $\mathcal{T}_1 \triangleq \{\emptyset\}$, and for every $t > 1$, $\mathcal{T}_t$ denotes the set of all possible transcripts of measurement outcomes $x_{<t}$ for which $x_i \in \Omega(\mathcal{M}^{x_{<i}})$ for all $1 \le i \le t-1$ (recall that $x_{<i} \triangleq (x_1, ..., x_{i-1})$). The schedule works in the natural manner: at time t for $t = 1, \dots, N$, given a transcript $x_{<t} \in \mathcal{T}_t$, it measures the t-th copy of ρ using the POVM $\mathcal{M}^{x_{<t}}$.*

If in addition every $\mathcal{M}^{x_{<t}}$ only depends on t and not on the specific transcript $x_{<t}$, we say it is a nonadaptive POVM schedule *and denote it simply by $\{\mathcal{M}^t\}_{t\in[N]}$.*

Definition 3.3 (Quantum property testing task)**.** *A* quantum property testing task *$\mathcal{T}$ is specified by two disjoint sets S_0 and S_1 of mixed states. For any $N \in \mathbb{N}$, we say that task T* has copy complexity *N if there exists a POVM schedule $\mathcal{S}$ and a (potentially randomized) post-processing algorithm A so that for any $\alpha \in \{0, 1\}$ and any $\rho \in S_\alpha$, if $z_{\le N}$ is the transcript obtained from measuring N copies of ρ according to $\mathcal{S}$, then $A(z_{\le N}) = \alpha$ with probability at least $2/3$ over the randomness of $\mathcal{S}$ and A.*

For a mixed state σ, if we specialize Definition 3.3 to $S_0 = \sigma$ and S_1 to all mixed states ε-far in trace distance from σ, we obtain the following standard task:

Definition 3.4 (State certification)**.** *Fix $\varepsilon > 0$. Given an explicit description of a mixed state σ along with copies of an unknown mixed state ρ, the task of* state certification to error ε with respect to σ *is to determine with high probability whether $\rho = \sigma$ or $\|\rho - \sigma\|_1 > \varepsilon$ by making measurements on the copies of ρ. When $\sigma = \rho_{\mathsf{mm}}$, this is the task of* mixedness testing.

We will employ the following standard framework for proving testing lower bounds:

Definition 3.5 (Lower Bound Setup: Point vs. Mixture)**.** *In the setting of Definition 3.3, a* point vs. mixture task *is specified by a null hypothesis $\rho \in S_0$, a set of alternatives $\rho_\theta \subseteq S_1$ parametrized by θ, and a distribution $\mathcal{D}$ over θ.*

For any POVM schedule $\mathcal{S}$, let $p_0^{\leq N}(\mathcal{S})$ be the induced distribution over transcripts from measuring N copies of ρ according to $\mathcal{S}$, and let $p_1^{\leq N}(\mathcal{S})$ be the induced distribution over transcripts from first sampling $\theta \sim \mathcal{D}$ and then measuring N copies of ρ_θ according to $\mathcal{S}$. For instance, if $\mathcal{S}$ is nonadaptive, then $p_1^{\leq N}$ is simply a mixture of product distributions.

The following is a standard fact that lets us relate this back to property testing:

Fact 3.6. *Given quantum property testing task $\mathcal{T}$ specified by sets S_0, S_1, let $N \in \mathbb{N}$, and let $\mathcal{F}$ be a family of measurement schedules using N measurements. Suppose there exists a point vs. mixture task so that for every $\mathcal{S} \in \mathcal{F}$, we have that $d_{TV}(p_0^{\leq N}(\mathcal{S}), p_1^{\leq N}(\mathcal{S})) \leq 1/3$. Then $\mathcal{T}$ has copy complexity at least N.*

For the remainder of the paper, we will fix a measurement schedule $\mathcal{S}$ and just write $p_0^{\leq N}$ and $p_1^{\leq N}$. The possible families $\mathcal{F}$ we work with in Fact 3.6 are the family of nonadaptive POVM schedules, and the family of adaptive POVM schedules.

3.2 Tail Bounds

We first collect some elementary facts about sub-exponential random variables.

Definition 3.7. *We say that a random variable Z is* (σ^2, b)-sub-exponential *if it has mean zero and satisfies*

$$\Pr[|Z| > s] \leq \exp\left(\frac{1}{2}\left\{\frac{s^2}{\sigma^2} \wedge \frac{s}{b}\right\}\right)$$

for all $s > 0$.

It is a standard fact that sub-exponential random variables satisfy the following moment bounds:

Lemma 3.8. *If Z is (σ^2, b)-sub-exponential, then for any $t \geq 1$, $\mathbb{E}[|Z|^t] \leq (t/2)! \cdot (2\sigma^2)^{t/2} + t! \cdot (2b)^t$.*

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[|Z|^t\right] &= \int_0^\infty \Pr\left[|Z| > s^{1/t}\right] ds \\
&\leq \int_0^\infty \exp\left(-\frac{s^{2/t}}{2\sigma^2}\right) ds + \int_0^\infty \exp\left(-\frac{s^{1/t}}{2b}\right) ds \\
&= \Gamma(1 + t/2) \cdot (2\sigma^2)^{t/2} + \Gamma(1 + t) \cdot (2b)^t
\end{aligned}$$

as desired. □

It is also standard that sub-exponential random variables have mgf bounded as follows:

Lemma 3.9. *If Z is (σ^2, b)-sub-exponential, then for any $\lambda \leq \min(1/4b, 1/\sigma)$,*

$$\mathbb{E}[e^{\lambda Z}] \leq \exp\left(O(\lambda^2(\sigma^2 + b^2))\right).$$

Proof. As $\mathbb{E}[Z] = 0$ by definition, we can expand

$$\mathbb{E}[e^{\lambda Z}] = 1 + \sum_{t=2}^{\infty} \frac{\lambda^t}{t!} \mathbb{E}[Z^t].$$

By Lemma 3.8,

$$\sum_{t=2}^{\infty} \frac{\lambda^t}{t!} \mathbb{E}[Z^t] \leq \sum_{t=2}^{\infty} \left(\frac{(t/2)!}{t!} (2\lambda^2\sigma^2)^{t/2} + (2\lambda b)^t \right) = 8\lambda^2 b^2 + \sum_{t=2}^{\infty} (\lambda^2\sigma^2/2)^{t/2} \leq 8\lambda^2 b^2 + \lambda^2\sigma^2.$$

The lemma follows from the inequality $1 + x \leq e^x$. □

We will need the following basic fact about sums of random variables satisfying sub-exponential moment bounds.

Lemma 3.10. *Fix any $t \in \mathbb{N}$. Given a collection of independent mean-zero random variables $Z_1, \ldots, Z_m$ whose odd moments vanish and such that for every $i \in [m]$ and even $1 \leq \ell \leq t$, $\mathbb{E}[|Z_i|^\ell]^{1/\ell} \leq \ell \cdot \sigma_i$, we have that for every even $1 \leq \ell \leq t$*

$$\mathbb{E}[(Z_1 + \cdots + Z_m)^\ell]^{1/\ell} \leq \ell(\sigma_1^2 + \cdots + \sigma_m^2)^{1/2}$$

Proof. Using the sub-exponential moment bound, we can expand $\mathbb{E}[(Z_1 + \cdots + Z_m)^\ell]$ and use the fact that the Z_i's are independent to get

$$\mathbb{E}[(Z_1 + \cdots + Z_m)^\ell] = \sum_{\alpha} \prod_i \mathbb{E}[Z_i^{\alpha_i}] \leq \sum_{\alpha} \prod_i \alpha_i^{\alpha_i} \sigma_i^{\alpha_i} \leq \ell^\ell \sum_{\alpha} \prod_i (\sigma_i^2)^{\alpha_i/2} = \ell^\ell (\sigma_1^2 + \cdots + \sigma_m^2)^{\ell/2}$$

where α ranges over even monomials of total degree ℓ. □

Concentration of measure for Haar-random unitary matrices will also be crucial to our analysis:

Theorem 3.11 ([MM13], Corollary 17, see also [AGZ10], Corollary 4.4.28)**.** *Equip $M \triangleq U(d)^k$ with the L_2-sum of Hilbert-Schmidt metrics. If $F : M \to \mathbb{R}$ is L-Lipschitz, then for any $t > 0$:*

$$\Pr_{(\mathbf{U}_1,\ldots,\mathbf{U}_k)\in M}[|F(\mathbf{U}_1, ..., \mathbf{U}_k) - \mathbb{E}[F(\mathbf{U}_1, ..., \mathbf{U}_k)]| \geq t] \leq e^{-dt^2/12L^2},$$

where $\mathbf{U}_1, ..., \mathbf{U}_k$ are independent unitary matrices drawn from the Haar measure.

3.3 Weingarten Calculus

In this section we recall some standard facts about integrals over the Haar measure on the unitary group. Given a permutation $\pi \in S_\ell$, let $\mathrm{Wg}(\pi, d)$ denote the *Weingarten function* (see e.g. [CŚ06]). Given a matrix $M \in \mathbb{C}^{d\times d}$ and permutation $\pi \in S_\ell$, let $\langle M \rangle_\pi \triangleq \prod_{C\in\pi} \mathrm{Tr}(M^{|C|})$, where C ranges over the cycles of π and $|C|$ denotes the length of C. Equivalently, if P_π is the permutation operator associated to π, then

$$\langle M \rangle_\pi = \mathrm{Tr}(P_\pi M^{\otimes\ell}). \tag{4}$$

We will use the following consequence of the Weingarten calculus and Schur-Weyl duality:

Lemma 3.12 (See e.g. Eq 7.32 from [BCHJ+19])**.** *For any matrix $\mathbf{M} \in (\mathbb{C}^{d\times d})^{\otimes\ell}$,*

$$\mathop{\mathbb{E}}_{\mathbf{U}}\left[\mathbf{U}^{\dagger\otimes\ell}\mathbf{M}\mathbf{U}^{\otimes\ell}\right] = \sum_{\sigma,\tau\in S_\ell} \mathrm{Wg}(\sigma^{-1}\tau, d)\,\mathrm{Tr}(P_\tau\mathbf{M})P_\sigma,$$

where the expectation is with respect to the Haar measure on $U(d)$.

Lemma 3.12 yields the following useful integral:

Lemma 3.13. *For $d \geq 2$, $\ell \in \mathbb{N}$, and any $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{d \times d}$, we have that*

$$\mathop{\mathbb{E}}_{\mathbf{U}}[\mathrm{Tr}(\mathbf{A}\mathbf{U}^\dagger \mathbf{B}\mathbf{U})^\ell] = \sum_{\pi,\tau \in S_\ell} \mathrm{Wg}(\pi^{-1}\tau, d)\langle \mathbf{A}\rangle_\pi \langle \mathbf{B}\rangle_\tau.$$

In particular, when $\ell = 1$, $\mathbb{E}_{\mathbf{U}}[\mathrm{Tr}(\mathbf{A}\mathbf{U}^\dagger\mathbf{B}\mathbf{U})] = \frac{1}{d}\mathrm{Tr}(\mathbf{A})\,\mathrm{Tr}(\mathbf{B})$.

Proof. We can write $\mathrm{Tr}(\mathbf{A}\mathbf{U}^\dagger\mathbf{B}\mathbf{U})^\ell$ as $\mathrm{Tr}\left(\mathbf{A}^{\otimes\ell}\mathbf{U}^{\dagger\otimes\ell}\mathbf{B}^{\otimes\ell}\mathbf{U}^{\otimes\ell}\right)$, so by Lemma 3.12, the expectation of this over $\mathbf{U}$ is $\sum_{\sigma,\tau\in S_\ell}\mathrm{Wg}(\sigma^{-1}\tau, d)\,\mathrm{Tr}(P_\tau \mathbf{B}^{\otimes\ell})\,\mathrm{Tr}(P_\sigma \mathbf{A}^{\otimes\ell})$, and the first part of the lemma then follows by (4). The second part of the lemma then follows by the fact that for the identity permutation e on one element, $\mathrm{Wg}(e, d) = \frac{1}{d}$. □

3.4 Block Matrices

Here we record two basic results about block matrices, beginning with the following standard fact about Schur complements (see e.g. Theorem 1.12 from [Zha06]):

Lemma 3.14 (Schur complements)**.** *For a block matrix $\rho = \begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^\dagger & \mathbf{C},\end{pmatrix}$ for which $\mathbf{A}$ and $\mathbf{C}$ are positive definite, ρ is positive definite if and only if Schur complement $\mathbf{C} - \mathbf{B}^\dagger\mathbf{A}^{-1}\mathbf{B}$ is positive definite.*

The second result of this subsection upper bounds the trace norm of the off-diagonal blocks of a psd block matrix in terms of the traces of the diagonal blocks:

Lemma 3.15. *For psd block matrix $\rho = \begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^\dagger & \mathbf{C},\end{pmatrix}$, where $\mathbf{A}$ and $\mathbf{C}$ are square, we have that $\mathrm{Tr}(\mathbf{A})\,\mathrm{Tr}(\mathbf{C}) \geq \|\mathbf{B}\|_1^2$. In particular, $\|\mathbf{B}\|_1 \leq \mathrm{Tr}(\rho)/2$.*

Proof. Without loss of generality suppose that $\mathbf{A}$ has at least as many rows/columns as $\mathbf{C}$. First note that we may assume $\mathbf{B}$ is actually square. Indeed, consider the matrix ρ' given by padding ρ with zeros,

$$\rho' = \begin{pmatrix}\mathbf{A} & \mathbf{B} & \mathbf{0}\\ \mathbf{B}^\dagger & \mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0},\end{pmatrix}$$

so that $\mathbf{A}$ and $\mathbf{C}' \triangleq \begin{pmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}$ have the same dimensions. Clearly, $\|\begin{pmatrix}\mathbf{B} & \mathbf{0}\end{pmatrix}\|_1 = \|\mathbf{B}\|_1$, and $\|\mathbf{C}'\|_1 = \|\mathbf{C}\|_1$, so to show Lemma 3.15 for ρ it suffices to prove it for ρ'. So henceforth, assume $\mathbf{B}$ is square.

We will further assume that $\mathbf{B}$ is diagonal. To see why this is without loss of generality, write the singular value decomposition $\mathbf{B} = \mathbf{U}^\dagger\mathbf{\Sigma}\mathbf{V}$ and note that

$$\begin{pmatrix}\mathbf{U} & \mathbf{0}\\ \mathbf{0} & \mathbf{V}\end{pmatrix}\rho\begin{pmatrix}\mathbf{U}^\dagger & \mathbf{0}\\ \mathbf{0} & \mathbf{V}^\dagger\end{pmatrix} = \begin{pmatrix}\mathbf{U}^\dagger\mathbf{A}\mathbf{U} & \mathbf{\Sigma}\\ \mathbf{\Sigma} & \mathbf{V}^\dagger\mathbf{C}\mathbf{V},\end{pmatrix}$$

If $\mathbf{B}$ is diagonal, then for every diagonal entry $\mathbf{B}_{i,i}$, we have that $\mathbf{B}_{i,i}^2 \leq \mathbf{A}_{i,i}\mathbf{C}_{i,i}$, so

$$\|\mathbf{B}\|_1^2 = \left(\sum_i \mathbf{B}_{i,i}\right)^2 \leq \left(\sum_i \mathbf{A}_{i,i}^{1/2}\mathbf{B}_{i,i}^{1/2}\right)^2 \leq \mathrm{Tr}(\mathbf{A})\,\mathrm{Tr}(\mathbf{B}),$$

where the last step is by Cauchy-Schwarz.

The second part of the claim follows by AM-GM. □

3.5 Instance-Optimal Distribution Testing

Here we record the precise statement of the instance-optimal lower bound from [VV17].

Theorem 3.16 ([VV17], Theorem 1)**.** *Given a known distribution p and samples from an unknown distribution q, any tester that can distinguish between $q = p$ and $\|p - q\|_1 \geq \varepsilon$ with probability 2/3 must draw at least $\Omega(1/\varepsilon \vee \|p_{-\varepsilon}^{-\max}\|_{2/3}/\varepsilon^2)$ samples.*

Note that this immediately implies a lower bound for state certification:

Corollary 3.17. *Given a known mixed state ρ and copies of an unknown mixed state σ, any tester that can distinguish between $\sigma = \rho$ and $\|\rho - \sigma\|_1 \geq \varepsilon$ with probability 2/3 using measurements on the copies of ρ must use at least $\Omega(\|\rho_{-\varepsilon}^{-\max}\|_{2/3}/\varepsilon^2)$ samples.*

We will use this corollary in our proof to handle mixed states whose eigenvalues are all pairwise separated by at least a constant factor. Intuitively, these mixed states are close to being low-rank, and one would expect that the copy complexity for testing identity to such a state is $\widetilde{\Theta}(1/\varepsilon^2)$. We show that this is indeed the case (see Lemma 5.12).

3.6 Miscellaneous Facts

The following elementary facts will be useful:

Fact 3.18. *Let S be any set of distinct positive integers. Given a collection of numbers $\{d_j\}_{j \in S}$ satisfying $\sum_j d_j 2^{-j} \leq 2$, let p be the vector with d_j entries equal to 2^{-j} for every $j \in S$. Then $\max_j d_j^b 2^{-aj} \geq |S|^{-b} \|p\|_{a/b}^{-a}$ for any $a, b > 0$.*

Proof. Let j^* be the index attaining the maximum. By maximality we know $d_{j^*} 2^{-aj/b} \geq \frac{1}{|S|} \sum_j d_j \cdot 2^{-aj/b}$. Raising both sides to the b-th power and taking reciprocals, we conclude that $2^{aj}/d_j^b \leq |S|^b \|p\|_{a/b}^a$. □

Fact 3.19. *Let $c > 1$ and $p, q > 0$. Given a vector v with entries $v_1 > \cdots > v_m > 0$ for which $v_i \geq c \cdot v_{i+1}$ for every i, we have that $\|v\|_p \geq (1 - c^{-q})^{1/q} \cdot \|v\|_q$.*

Proof. We have that $\|v\|_q^q \leq \sum_{i=1}^{\infty} (c^{-i} v_1)^q = \frac{v_1^q}{1 - c^{-q}}$, so $\|v\|_p \geq v_1 \geq \|v\|_q \cdot (1 - c^{-q})^{1/q}$. □

We will also need the following when describing the framework of [BCL20] in Section 4:

Fact 3.20 (Integration by parts, see e.g. Fact C.2 in [BCL20])**.** *Let $a, b \in \mathbb{R}$. Let Z be a nonnegative random variable satisfying $Z \leq b$ and such that for all $x \geq a$, $\Pr[Z > x] \leq \tau(x)$. Let $f : [0, b] \to \mathbb{R}_{\geq 0}$ be nondecreasing and differentiable. Then*

$$\mathbb{E}[f(Z)] \leq f(a)(1 + \tau(a)) + \int_a^b \tau(x) f'(x) \, dx.$$

4 General Lower Bound Framework

All of our lower bounds are based on analyzing a suitable point vs. mixture distinguishing problem. In this section we outline a general framework, implicit in [BCL20], for showing copy complexity lower bounds for such problems. After outlining some basic objects, in Section 4.1 we describe a set of conditions (see Assumption 1) that, if true for a particular distinguishing problem, imply by the machinery of [BCL20] a strong copy complexity lower bound for that problem. We formally state

these implications in Sections 4.2 and 4.3 and, for the sake of completeness, provide their proofs in Appendix B.2.[3]

Concretely, we will lower bound the smallest N for which it is possible to distinguish, using an unentangled POVM schedule $\mathcal{S}$, between $\sigma^{\otimes N}$ and $\mathbb{E}_{\mathbf{U}\sim\mathcal{D}}[\rho_{\mathbf{U}}^{\otimes N}]$ for some prior distribution $\mathcal{D}$. Given schedule $\mathcal{S}$, let $p_0^{\leq N}$ (resp. $p_1^{\leq N}$) denote the distribution over transcripts given by measuring $\sigma^{\otimes N}$ (resp. $\mathbb{E}_{\mathbf{U}}[\rho_{\mathbf{U}}^{\otimes N}]$) with $\mathcal{S}$. A key component of our analysis is to bound how well a single step of $\mathcal{S}$ can distinguish between a single copy of σ and a single copy of $\sigma_{\mathbf{U}}$ for $\mathbf{U}\sim\mathcal{D}$:

Definition 4.1. *A* single-copy sub-problem $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ *consists of the following data: a POVM* $\mathcal{M}$ *over* $\mathbb{C}^d$*, a mixed state* $\sigma \in \mathbb{C}^{d\times d}$*, and a distribution over mixed states* $\sigma_{\mathbf{U}} \in \mathbb{C}^{d\times d}$ *where* $\mathbf{U}$ *is drawn from some distribution* $\mathcal{D}$*.*

To quantify how much information a single step of $\mathcal{S}$ can reveal about the unknown state, we introduce the following quantities:

Definition 4.2. *Given a single-copy sub-problem* $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$*, let* $p_0(\mathcal{M})$ *denote the distribution over outcomes upon measuring* σ *using* $\mathcal{M} = \{M_z\}$*. Given POVM outcome* z*, and* $\mathbf{U}, \mathbf{V} \in$ *supp*$(\mathcal{D})$*, define the quantities*

$$g_{\mathcal{P}}^{\mathbf{U}}(z) \triangleq \frac{\langle M_z, \sigma_{\mathbf{U}}\rangle}{\langle M_z, \sigma\rangle} - 1 \qquad \phi_{\mathcal{P}}^{\mathbf{U},\mathbf{V}} \triangleq \mathop{\mathbb{E}}_{z\sim p_0(\mathcal{M})}\left[g_{\mathcal{P}}^{\mathbf{U}}(z)\cdot g_{\mathcal{P}}^{\mathbf{V}}(z)\right].$$

We will omit the subscript $\mathcal{P}$ *when the context is clear.*

We can interpret $1 + g_{\mathcal{P}}^{\mathbf{U}}$ as the likelihood ratio between the distribution under measuring a single copy of $\sigma_{\mathbf{U}}$ and the distribution under measuring a single copy of σ.

4.1 Sufficient Conditions on $g_{\mathcal{P}}^{\mathbf{U}}(z)$

We will design $\{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}}$ in such a way that the following three conditions hold.

Assumption 1. *Suppose that* $g_{\mathcal{P}}^{\mathbf{U}}$ *satisfies the following three properties for parameters* $\varsigma, L > 0$*:*

1. *First moment bound: For any* $z \in \Omega(\mathcal{M})$*,* $\mathbb{E}_{\mathbf{U}}[g_{\mathcal{P}}^{\mathbf{U}}(z)] = 0$*.*
2. *Second moment bound:* $\mathbb{E}_{\mathbf{U}\sim\mathcal{D}}[g_{\mathcal{P}}^{\mathbf{U}}(z)^2] \leq \varsigma^2$ *for all measurement outcomes* z*.*
3. *Lipschitzness:* $\mathbb{E}_{z\sim p_0(\mathcal{M})}[(g_{\mathcal{P}}^{\mathbf{U}}(z) - g_{\mathcal{P}}^{\mathbf{V}}(z))^2]^{1/2} \leq L\cdot\|\mathbf{U}-\mathbf{V}\|_{HS}$ *for any* $\mathbf{U}, \mathbf{V} \in$ *supp*$(\mathcal{D})$*.*

Example 4.3. *It was shown in [BCL20] that if* $\sigma = \rho_{\mathsf{mm}}$*,* $\sigma_{\mathbf{U}} = \rho_{\mathsf{mm}} + \mathbf{U}^\dagger diag(\frac{\varepsilon}{d}, \ldots, -\frac{\varepsilon}{d}, \ldots)\mathbf{U}$*, and* $\mathcal{D}$ *is given by the Haar measure over* $U(d)$*, then Assumption 1 holds for* $\varsigma, L = O(\varepsilon/\sqrt{d})$ *for any sub-problem* $\mathcal{P}$ *of the form* $(\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$*.*

Here we prove some intuition for these conditions. As we mentioned above, $1 + g_{\mathcal{P}}^{\mathbf{U}}$ is simply the likelihood ratio between the distributions over outcomes under measuring a single copy of $\sigma_{\mathbf{U}}$ versus a single copy of σ. Condition 1 thus ensures that for any POVM element z, the probability of observing outcome z under $\sigma_{\mathbf{U}}$ is in expectation over $\mathbf{U}$ equal to the probability of observing z under σ. By Chebyshev's, Condition 2 then ensures that the former has some mild concentration around the latter.

[3] That said, as our techniques are a generalization of the approach of [BCL20], readers unfamiliar with that work may find it more convenient to consult it first before proceeding. Either way, here we will try to distill the main ingredients from [BCL20] in as modular a fashion as possible.

In other words, because of Conditions 1 and 2, there is no single observable that we can repeatedly measure $O(1/\varsigma^2)$ times to solve the point vs. mixture distinguishing problem. It turns out that if $g_{\mathcal{P}}^{\mathbf{U}}$ additionally satisfies the Lipschitzness constraint of Condition 3, then we can invoke concentration of Lipschitz functions of Haar-random unitary matrices (recall Theorem 3.11 from the preliminaries) to get a strong lower bound for the distinguishing problem.

This last point requires some unpacking. For starters, let us spell out what kinds of tail bounds we leverage. Specifically, using Assumption 1 and concentration of measure, one can show the following tail bound which is an important starting point for our lower bounds.

Lemma 4.4. *Suppose $\mathcal{P}$ satisfies Assumption 1 for parameters $\varsigma, L > 0$. Then for $\mathbf{U}, \mathbf{V}$ sampled independently from the Haar measure over $U(d)$, $\phi_{\mathcal{P}}^{\mathbf{U},\mathbf{V}}$ is a $\left(\Theta(\varsigma^2 L^2/d), \Theta(L^2/d)\right)$-sub-exponential random variable in the randomness of $\mathbf{U}, \mathbf{V}$. In particular, by Lemma 3.8,*

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\left|\phi_{\mathcal{P}}^{\mathbf{U},\mathbf{V}}\right|^t\right]^{1/t} \leq O\left(\varsigma L\sqrt{t/d} \vee L^2 t/d\right) \leq O(t \cdot L \cdot \{\varsigma \vee L\}/\sqrt{d}) \tag{5}$$

In the next two sections, we show how to use Lemma 4.4 to derive lower bounds for the distinguishing problem.

4.2 Non-adaptive Lower Bounds

As discussed in Section 2, our non-adaptive lower bounds are based on the Ingster-Suslina method [IS12]. In [BCL20], the main ingredients of this method are stated in the preceding notation as follows:

Lemma 4.5 ([BCL20], Lemma 2.8)**.** *If the unentangled POVM schedule $\mathcal{S}$ is non-adaptive and consists of POVMs $\mathcal{M}_1, ..., \mathcal{M}_N$, then if $\mathcal{P}_t = (\mathcal{M}_t, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ denotes the t-th single-copy sub-problem for an arbitrary $\mathcal{D}$, then*

$$\chi^2\left(p_1^{\leq N} \| p_0^{\leq N}\right) \leq \max_{t\in[N]} \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}\sim\mathcal{D}}\left[\left(1 + \phi_{\mathcal{P}_t}^{\mathbf{U},\mathbf{V}}\right)^N\right] - 1 \tag{6}$$

Lemma 4.5 is one reason why we care about tail bounds for $\phi_{\mathcal{P}}^{\mathbf{U},\mathbf{V}}$: with sufficiently good moment bounds on ϕ, we can upper bound the right-hand side of (6) and conclude that for N small, the chi-squared divergence between $p_1^{\leq N}$ and $p_0^{\leq N}$ is small. By Pinsker's, this implies that the total variation distance between $p_1^{\leq N}$ and $p_0^{\leq N}$ is small, so by Fact 3.6 we get a lower bound on the copy complexity N of distinguishing $\sigma^{\otimes N}$ and $\mathbb{E}[\sigma_{\mathbf{U}}^{\otimes N}]$. We spell this out explicitly in the next lemma.

Lemma 4.6. *Let $\mathcal{D}$ be the Haar measure over $U(d)$, and fix σ and $\{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}}$. Suppose that for any POVM $\mathcal{M}$, the single-copy sub-problem $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ satisfies Assumption 1. Then distinguishing $\sigma^{\otimes N}$ from $\mathbb{E}_{\mathbf{U}}[\rho_{\mathbf{U}}^{\otimes N}]$ with probability at least 2/3 using an unentangled, non-adaptive POVM schedule $\mathcal{S}$ requires $N = \Omega\left(\sqrt{d}/(L\varsigma) \wedge d/L^2\right)$.*

Proof. Fix any $t \in [N]$ and note that $(1+\phi_{\mathcal{P}_t}^{\mathbf{U},\mathbf{V}})^N \leq \exp\left(N\phi_{\mathcal{P}_t}^{\mathbf{U},\mathbf{V}}\right)$. As $\phi_{\mathcal{P}_t}^{\mathbf{U},\mathbf{V}}$ is $\left(\Theta(\varsigma^2 L^2/d), \Theta(L^2/d)\right)$-sub-exponential, its moment generating function is bounded by Lemma 3.9. In particular, for any $N \leq O(d/L^2)$,

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\exp\left(N\phi_{\mathcal{P}_t}^{\mathbf{U},\mathbf{V}}\right)\right] \leq \exp\left(O(N^2(\varsigma^2 L^2/d + L^4/d^2)\right),$$

so for $N = o\left(\sqrt{d}/(L\varsigma) \wedge d/L^2\right)$, the above quantity is $1 + o(1)$. The lemma then follows from relating KL to total variation using Pinsker's and then invoking Fact 3.6. □

Example 4.7. *If $\sigma = \rho_{\mathsf{mm}}$, $\sigma_{\mathbf{U}} = \rho_{\mathsf{mm}} + \mathbf{U}^\dagger diag(\frac{\varepsilon}{d}, \ldots, -\frac{\varepsilon}{d}, \ldots)\mathbf{U}$, and $\mathcal{D}$ is the Haar measure on $U(d)$, recall from Example 4.3 that we can take $\varsigma, L = O(\varepsilon/\sqrt{d})$. So by Lemma 4.6 we get a lower bound of $N = \Omega(d^{3/2}/\varepsilon^2)$. This recovers the non-adaptive lower bound for mixedness testing from [BCL20].*

4.3 Adaptive Lower Bounds

For our adaptive lower bounds, we follow the chain rule-based framework introduced in [BCL20], the main result of which can be abstracted as follows:

Theorem 4.8 (Implicit in [BCL20])**.** *Let $\mathcal{D}$ be the Haar measure over $U(d)$, and fix σ and $\{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}}$. Suppose that for any POVM $\mathcal{M}$, the single-copy sub-problem $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ satisfies Assumption 1 and additionally, for all $z \in \Omega(\mathcal{M})$, $|g_{\mathcal{P}}^{\mathbf{U}}(z)| \le 0.99$ almost surely. Then for any $\tau > 0$ and $N = o(d/L^2)$,*

$$KL\left(p_1^{\le N} \| p_0^{\le N}\right) \le N\tau + O(N) \cdot \exp\left(-\Omega\left(\left\{\frac{d\tau^2}{L^2\varsigma^2} \wedge \frac{d\tau}{L^2}\right\} - N \cdot \varsigma^2\right)\right). \tag{7}$$

Like the proof of Theorem 4.6, the proof of Theorem 4.8 also makes crucial use of the fact that $\phi_{\mathcal{P}}^{\mathbf{U},\mathbf{V}}$ is a sub-exponential random variable. As it is somewhat more involved, we defer the proof to Appendix B.2.

Example 4.9. *Take any $\varepsilon \le 0.99$. If $\sigma = \rho_{\mathsf{mm}}$ and $\sigma_{\mathbf{U}} = \rho_{\mathsf{mm}} + \mathbf{U}^\dagger diag(\frac{\varepsilon}{d}, \ldots, -\frac{\varepsilon}{d}, \ldots)\mathbf{U}$ as in Example 4.3, where recall that $\mathbf{U} \sim \mathcal{D}$ for $\mathcal{D}$ given by the Haar measure over $U(d)$, then note that*

$$|g_{\mathcal{P}}^{\mathbf{U}}(z)| \le \|\mathbf{U} diag(\varepsilon, \ldots, -\varepsilon, \ldots)\mathbf{U}^\dagger\| = \varepsilon \le 0.99$$

for any sub-problem $\mathcal{P}$ of the form $(\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$. So by taking $\tau = \varepsilon^2/d^{4/3}$ in Theorem 4.8, one gets that for $N = o(d^{4/3}/\varepsilon^2)$, the KL divergence in (7) is $o(1)$. This recovers the $\Omega(d^{4/3}/\varepsilon^2)$ adaptive lower bound for mixedness testing from [BCL20].

5 Nonadaptive Lower Bound for State Certification

In this section we will show our instance-near-optimal lower bounds for state certification with nonadaptive, unentangled measurements.

Theorem 5.1. *There is an absolute constant $c > 0$ for which the following holds for any $0 < \varepsilon < c$.*[4] *Let $\sigma \in \mathbb{C}^{d\times d}$ be a diagonal density matrix. There is a matrix σ^{**} given by zeroing out at most $O(\varepsilon)$ mass from σ (see Definition 5.2 and Fact 5.3 below), such that the following holds:*

Let $\widehat{\sigma}^{**} \triangleq \sigma^{**}/\operatorname{Tr}(\sigma^{**})$, and let d_{eff} denote the number of nonzero entries of σ^{**}. Then any algorithm for state certification to error ε with respect to σ using nonadaptive, unentangled measurements has copy complexity at least

$$\Omega\left(d\sqrt{d_{\mathsf{eff}}} \cdot F(\widehat{\sigma}^{**}, \rho_{\mathsf{mm}})/(\varepsilon^2 \operatorname{polylog}(d/\varepsilon))\right).$$

In Section 5.1, we describe a bucketing scheme that will be essential to our analysis. In Section 5.2 we describe and analyze the first of our two lower bound instances, a distinguishing problem

[4] As presented, our analysis yields c within the vicinity of 1/3, but we made no attempt to optimize for this constant.

based on a generalization of the standard quantum Paninski construction. Specifically, in Section 5.2.1, we give a generic copy complexity lower bound for this problem, and in Section 5.2.2 we show how to tune the relevant parameters to obtain a copy complexity lower bound based on the Schatten 2/5-quasinorm of σ. In Section 5.3, we describe and analyze the second of our two lower bound instances, a distinguishing problem based on perturbing the off-diagonal entries of an appropriately chosen principal submatrix of σ, obtaining for restricted choices of ε a copy complexity lower bound based on the effective dimension and Schatten 1/2-quasinorm of σ. In Section 5.5, we put together the analyses of our two lower bound instances to conclude the proof of Theorem 5.1.

5.1 Bucketing and Mass Removal

We may without loss of generality assume that σ is some diagonal matrix $\mathrm{diag}(\lambda_1, \dots, \lambda_d)$.

For $j \in \mathbb{Z}_{\geq 0}$, let S_j denote the set of indices $i \in [d]$ for which $\lambda_i \in [2^{-j-1}, 2^{-j}]$; denote $|S_j|$ by d_j. Let $\mathcal{J}$ denote the set of j for which $S_j \neq \emptyset$. We will refer to $j \in \mathcal{J}$ as *buckets*. It will be convenient to refer to the index of the bucket containing a particular index $i \in [d]$ as $j(i)$. Also let S_{sing} denote the set of $i \in [d]$ belonging to a size-1 bucket S_j for some $j \in \mathcal{J}$, and let S_{many} denote the set of $i \in [d]$ which lie in a bucket S_j of size greater than 1 for some $j \in \mathcal{J}$.

Our bounds are based on the following modification of σ obtained by zeroing out a small fraction of its entries:

Definition 5.2 (Removing low-probability elements- nonadaptive lower bound)**.** *Without loss of generality, suppose that $\lambda_1, \dots, \lambda_d$ are sorted in ascending order according to $\lambda_i / d_{j(i)}^2$.*[5] *Let $d' \leq d$ denote the largest index for which $\sum_{i=1}^{d'} \lambda_i \leq 3\varepsilon$. Let $S_{\mathsf{tail}} \triangleq [d']$, and let S_{light} be the set of $i \in \{d'+1, \dots, d\}$ for which $\sum_{i' \in S_{j(i)} \setminus S_{\mathsf{tail}}} \lambda_{i'} \leq 2\varepsilon / \log(d/\varepsilon)$.*

Let $i_{\max}$ denote the index of the largest entry of σ. Let σ' denote the matrix given by zeroing out the largest entry of σ and the entries indexed by S_{tail}, and let σ^* denote the matrix given by zeroing out the entries indexed by $S_{\mathsf{tail}} \cup S_{\mathsf{light}}$. Finally, let σ^{**} denote the matrix given by further zeroing out from σ^* as many of the smallest entries as possible without removing more than 2ε mass.

Lastly, it will be convenient to define $\mathcal{J}'$ (resp. $\mathcal{J}^*$) to be the set of $j \in \mathcal{J}$ for which S_j has nonempty intersection with $(([d] \setminus \{i_{\max}\}) \cap S_{\mathsf{many}}) \setminus S_{\mathsf{tail}}$ (resp. $[d] \setminus (S_{\mathsf{tail}} \cup S_{\mathsf{light}})$). Note that by design, $\mathcal{J}'$ and $\mathcal{J}^*$ denote the indices of the nonzero diagonal entries of σ' and σ^* respectively.

We will use the following basic consequence of bucketing:

Fact 5.3. *There are at most $O(\log(d/\varepsilon))$ indices $j \in \mathcal{J}$ for which S_j and S_{tail} are disjoint. As a consequence, $\mathrm{Tr}(\sigma^{**}) \geq 1 - O(\varepsilon)$.*

Proof. For any $i_1 \notin S_{\mathsf{tail}}$ and $i_2 \in S_{\mathsf{tail}}$, we have that $p_{i_1} / d_{j(i_1)}^2 \geq p_{i_2} / d_{j(i_2)}^2$, so $p_{i_1} \geq p_{i_2} / d^2$. In particular, summing over $i_2 \in S_{\mathsf{tail}}$, we conclude that $p_{i_1} \cdot |S_{\mathsf{tail}}| \geq \varepsilon / d^2$, so $p_{i_1} \geq \varepsilon / d^3$. By construction of the buckets S_j, the first part of the claim follows. For the second part, by definition we have that $\sum_{i \in [d']} \lambda_i \leq O(\varepsilon)$. Furthermore, $\sum_{i \in S_{\mathsf{light}}} \lambda_i = O(\varepsilon)$ because of the first part of the claim. The second part of the claim follows by triangle inequality. □

Lastly, we will use the following shorthand: for any $j \in \mathcal{J}$ and any matrix $\mathbf{A}$, we will let $\mathbf{A}_j \in \mathbb{R}^{d \times d}$ denote the matrix which is zero outside of the principal submatrix indexed by S_j and which agrees with $\mathbf{A}$ within this submatrix.

[5] The only place where we need this particular choice of sorting is in the proof of Corollary 5.17 below.

5.2 Lower Bound Instance I: General Quantum Paninski

We will analyze the following distinguishing problem. We will pick a diagonal matrix $\boldsymbol{\mathcal{E}}$ as follows:

Definition 5.4 (Perturbation matrix $\boldsymbol{\mathcal{E}}$)**.** *For any $i \notin S_{\mathsf{many}}$, we will take the i-th diagonal entry of $\boldsymbol{\mathcal{E}}$ to be zero. For any bucket j of size at least 2, we will take the nonzero diagonal entries of $\boldsymbol{\mathcal{E}}_j$ to be $(\varepsilon_j, \cdots, -\varepsilon_j, \cdots)$ where there are $\lfloor d_j/2 \rfloor$ copies of ε_j and $\lfloor d_j/2 \rfloor$ copies of $-\varepsilon_j$, for ε_j to be optimized later.*

Given $\mathbf{U} \in U(d)$, define $\sigma_{\mathbf{U}} \triangleq \sigma + \mathbf{U}^\dagger \boldsymbol{\mathcal{E}} \mathbf{U}$.

Throughout this subsection, let $\mathcal{D}$ denote the distribution over block-diagonal unitary matrices $\mathbf{U}$ which are zero outside of the principal submatrices indexed by S_j for some $j \in \mathcal{J}$ with $d_j > 1$, and which within each submatrix indexed by such an S_j is an independent Haar-random unitary if d_j is even, and otherwise is an independent Haar-random unitary in the submatrix consisting of the first $2\lfloor d_j/2 \rfloor$ rows/columns. This distinction will not be particularly important in the sequel, so the reader is encouraged to imagine that d_j is always even when $d_j > 1$.

The objective of this subsection is to show the following lower bound:

Lemma 5.5. *Fix $0 < \varepsilon < c$ for sufficiently small absolute constant $c > 0$. Let $\sigma \in \mathbb{C}^{d \times d}$ be a diagonal density matrix. There is a choice of $\boldsymbol{\mathcal{E}}$ in Definition 5.4 for which distinguishing between whether $\rho = \sigma$ or whether $\rho = \sigma + \mathbf{U}^\dagger \boldsymbol{\mathcal{E}} \mathbf{U}$ for $\mathbf{U} \sim \mathcal{D}$ using nonadaptive, unentangled measurements has copy complexity at least $\Omega(\|\sigma'\|_{2/5}/(\varepsilon^2 \log(d/\varepsilon)))$.*

By definition of $\mathcal{D}$, ρ is block-diagonal in either scenario, and the block-diagonal structure depends only on $\{S_j\}$. In particular, this implies that we can without loss of generality assume that the POVMs the tester uses respect this block structure. More precisely:

Lemma 5.6. *Let $\rho \in \mathbb{C}^{d \times d}$ be any density matrix which is zero outside of the principal submatrices indexed by the subsets $\{S_j\}_{j \in \mathcal{J}}$. Given an arbitrary POVM $\mathcal{M} = \{M_z\}$, there is a corresponding POVM $\mathcal{M}'$ satisfying the following. Let p, p' be the distributions over measurement outcomes from measuring ρ with $\mathcal{M}, \mathcal{M}'$ respectively. Then:*

- *For every $z \in \Omega(\mathcal{M}')$, there exists $j \in \mathcal{J}$ for which M'_z is zero outside of the principal submatrix indexed by S_j*
- *There is a function $f : \Omega(\mathcal{M}') \to \Omega(\mathcal{M})$ for which the pushforward of p' under f is p.*

Proof. For every $z \in \Omega(\mathcal{M})$ and every $j \in \mathcal{J}$, define a POVM element $M_{j,z} \triangleq \Pi_j M_z \Pi_j$, where $\Pi_j \in \mathbb{C}^{d \times d}$ is the matrix which is equal to the identity in the principal submatrix indexed by S_j and is zero elsewhere. Clearly $\{M_{j,z}\}_{j \in \mathcal{J}, z \in \Omega(\mathcal{M})}$ is still a POVM because $\sum \Pi_j = \mathbb{1}$; let $\mathcal{M}'$ be this POVM. Let f be given by $f((j,z)) = z$. The pushforward of p' under f places mass

$$\sum_{j \in \mathcal{J}} \langle \rho, \Pi_j M_z \Pi_j \rangle = \left\langle \sum_{j \in \mathcal{J}} \Pi_j \rho \Pi_j, M_z \right\rangle = \langle \rho, M_z \rangle$$

on $z \in \Omega(\mathcal{M})$ as claimed, where the penultimate step follows by the assumption that ρ is zero outside of the principal submatrices indexed by the subsets $\{S_j\}$. □

By Lemma 5.6, we will henceforth only work with POVMs like $\mathcal{M}'$. If $\mathcal{M}^t$ is the t-th POVM used by the tester, we may assume without loss of generality that its outcomes $\Omega(\mathcal{M}^t)$ consist of pairs (j, z), where the POVM element corresponding to such a pair has nonzero entries in the principal submatrix indexed by S_j. Henceforth, fix an arbitrary such POVM $\mathcal{M}$ (we will drop

subscripts accordingly) and denote its elements by $\{M_{j,z}\}$ for $j \in \mathcal{J}$. We will denote by Ω_j the set of z for which there is an element $M_{j,z}$.

Let p denote the distribution over $\mathcal{J}$ induced by measuring σ with $\mathcal{M}$ and recording which bucket the outcome belongs to. Concretely, p places mass $p_j \triangleq \sum_{z\in\Omega_j}\langle M_{j,z}, \sigma_j\rangle = \mathrm{Tr}(\sigma_j)$ on bucket $j \in \mathcal{J}$. Similarly, define q^j to be the distribution over Ω_j conditioned on the outcome falling in bucket j, that is, q^j places mass $q^j_z \triangleq \frac{1}{p_j}\langle M_{j,z}, \sigma_j\rangle$ on $z \in \Omega_j$.

For every $j \in \mathcal{J}$, let $\mathcal{P}_j$ denote the single-copy sub-problem in d_j dimensions given by restricting to the coordinates indexed by S_j and using the POVM $\mathcal{M}_j \triangleq \{(M_{j,z})_j\}_{z\in\Omega_j}$. Formally, $\mathcal{P}_j$ is specified by the data $(\mathcal{M}_j, \widehat{\sigma}_j, \{(\widehat{\sigma}_{\mathbf{U}})_j\}_{\mathbf{U}\sim\mathcal{D}_j})$, where $\mathcal{D}_j$ is the Haar measure over $U(d_j)$ if d_j is even and is otherwise the distribution over $d_j \times d_j$ matrices which are Haar-random unitary in the first $2\lfloor d_j/2\rfloor$ rows/columns and zero elsewhere. Note that the density matrix $(\widehat{\sigma}_{\mathbf{U}})_j$ can be written as $\widehat{\sigma}_j + \mathbf{U}^\dagger \boldsymbol{\mathcal{E}}'_j \mathbf{U}$ for $\boldsymbol{\mathcal{E}}'_j \triangleq \boldsymbol{\mathcal{E}}_j/p_j$.

For any $j \in \mathcal{J}$, $z \in \Omega_j$, it will be convenient to define $\widetilde{M}_{j,z} \triangleq \frac{1}{\langle M_{j,z},\sigma_j\rangle} M_{j,z}$. We can write

$$g^{\mathbf{U}_j}_{\mathcal{P}_j}(z) = \frac{\langle M_{j,z}, \mathbf{U}_j^\dagger \boldsymbol{\mathcal{E}}'_j \mathbf{U}_j\rangle}{\langle M_{j,z}, \widehat{\sigma}_j\rangle} = \frac{\langle M_{j,z}, \mathbf{U}_j^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U}_j\rangle}{\langle M_{j,z}, \sigma_j\rangle} = \langle \widetilde{M}_{j,z}, \mathbf{U}_j^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U}_j\rangle. \tag{8}$$

Because $M_{j,z}$ is zero outside of the principal submatrix indexed by S_j, we thus have

$$g^{\mathbf{U}}(z) = \frac{\langle M_{j,z}, \mathbf{U}^\dagger \boldsymbol{\mathcal{E}} \mathbf{U}\rangle}{\langle M_{j,z}, \sigma\rangle} = \frac{\langle M_{j,z}, \mathbf{U}_j^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U}_j\rangle}{\langle M_{j,z}, \sigma_j\rangle} = g^{\mathbf{U}_j}_{\mathcal{P}_j}(z)$$

and

$$\phi^{\mathbf{U},\mathbf{V}} = \mathop{\mathbb{E}}_{j,z}\left[\frac{\langle M_{j,z}, \mathbf{U}_j^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U}_j\rangle\langle M_{j,z}, \mathbf{V}_j^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{V}_j\rangle}{\langle M_{j,z}, \sigma_j\rangle^2}\right] = \sum_{j\in\mathcal{J}} p_j \cdot \phi^{\mathbf{U}_j,\mathbf{V}_j}_{\mathcal{P}_j}. \tag{9}$$

We now give a generic lower bound for the distinguishing problem in Lemma 5.5 that depends on the entries of $\boldsymbol{\mathcal{E}}$. After that, we show how to tune the entries of $\boldsymbol{\mathcal{E}}$ to complete the proof of Lemma 5.5.

5.2.1 Bound Under General Perturbations

Our goal is first to show the following generic bound:

Lemma 5.7. *Distinguishing $\sigma^{\otimes N}$ from $\mathbb{E}_{\mathbf{U}}[\sigma_{\mathbf{U}}^{\otimes N}]$ with probability at least 2/3 using an unentangled, adaptive POVM schedule $\mathcal{S}$ requires*

$$N = \Omega\left(\left(\sum_{j\in\mathcal{J}} \frac{2^{2j}\varepsilon_j^4}{d_j}\right)^{-1/2}\right) \tag{10}$$

By Lemma 4.5, it suffices to show that for any POVM $\mathcal{M}$, $\mathbb{E}_{\mathbf{U},\mathbf{V}}\left[\left(1+\phi^{\mathbf{U},\mathbf{V}}_{\mathcal{M}}\right)^N\right] = 1+o(1)$ for N smaller than the claimed bound. To do this, we will bound the moments of each $\phi^{\mathbf{U},\mathbf{V}}_{\mathcal{P}_j}$ individually.

As the relevant matrices $(M_{j,z})_j$ are zero outside of the principal submatrix indexed by S_j, we will abuse notation and refer to them as $M_{j,z}$ in the sequel whenever the context is clear. Likewise, we will refer to $\mathbf{U}_j \sim \mathcal{D}_j$ as $\mathbf{U}$.

In the next three lemmas, we verify that the three conditions of Assumption 1 are satisfied for appropriate choices of ς, L by the d_j-dimensional single-copy sub-problem $\mathcal{P}_j$. For the proofs of these lemmas, it will be convenient to define $\widetilde{M}_{j,z} \triangleq \frac{1}{\langle M_{j,z},\sigma_j\rangle} M_{j,z}$

Lemma 5.8. *For any $z \in \Omega_j$, $\mathbb{E}_{\mathbf{U}}[g^{\mathbf{U}}_{\mathcal{P}_j}(z)] = 0$, so Condition 1 of Assumption 1 holds.*

Proof. By the second part of Lemma 3.13, $\mathbb{E}_{\mathbf{U}}[g^{\mathbf{U}}_{\mathcal{P}_j}(z)] = \mathrm{Tr}(\widetilde{M}_{j,z}) \cdot \mathrm{Tr}(\boldsymbol{\mathcal{E}}_j) = 0$. □

Lemma 5.9. $\mathbb{E}_{\mathbf{U}}[g^{\mathbf{U}}_{\mathcal{P}_j}(z)^2]^{1/2} \le O(2^j\varepsilon_j/\sqrt{d_j})$ *for any $z \in \Omega_j$, so Condition 2 of Assumption 1 holds.*

Proof. Let $\tau^* \in S_2$ denote transposition. For any $z \in \Omega_j$, by (8) and Lemma 3.13,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{U}}[g^{\mathbf{U}}_{\mathcal{P}_j}(z)^2] &= \mathbb{E}\left[\left\langle \widetilde{M}_{j,z}, \mathbf{U}^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U}\right\rangle^2\right] \\
&= \sum_{\pi,\tau\in S_2} \langle \boldsymbol{\mathcal{E}}_j\rangle_\tau \langle \widetilde{M}_{j,z}\rangle_\pi \,\mathrm{Wg}(\pi\tau^{-1}, d_j) \\
&= \langle \boldsymbol{\mathcal{E}}_j\rangle_{\tau^*}\left(\mathrm{Tr}(\widetilde{M}^2_{j,z}) \cdot \mathrm{Wg}(e, d_j) + \mathrm{Tr}(\widetilde{M}_{j,z})^2 \cdot \mathrm{Wg}(\tau^*, d_j)\right) \\
&\le d_j \cdot \varepsilon_j^2 \cdot \frac{\mathrm{Tr}(M_{j,z})^2}{\langle M_{j,z}, \sigma_j\rangle^2}\left(\frac{1}{d_j^2-1}\mathrm{Tr}(\widehat{M}^2_{j,z}) - \frac{1}{d_j(d_j^2-1)} \cdot \mathrm{Tr}(\widehat{M}_{j,z})^2\right) \\
&\le \frac{\varepsilon_j^2}{d_j+1} \cdot \frac{\mathrm{Tr}(M_{j,z})^2}{\langle M_{j,z}, \sigma_j\rangle^2} \le 2 \cdot 2^{2j}\varepsilon_j^2/d_j,
\end{aligned}$$

where in the last step we used the fact that $\mathrm{Tr}(\widehat{M^2}) \le 1$ for any matrix $\widehat{M}$ of trace 1. □

Lemma 5.10. $\mathbb{E}_{z\sim q^j}[(g^{\mathbf{U}}_{\mathcal{P}_j}(z) - g^{\mathbf{V}}_{\mathcal{P}_j}(z))^2]^{1/2} \le O((2^j/p_j)^{1/2}\varepsilon_j) \cdot \|\mathbf{U}-\mathbf{V}\|_{HS}$ *for any $\mathbf{U}, \mathbf{V} \in U(d)$, so Condition 3 of Assumption 1 holds.*

Proof. The matrix $\mathbf{A} \triangleq \mathbf{U}^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U} - \mathbf{U}'^\dagger \boldsymbol{\mathcal{E}}_j \mathbf{U}'$ is Hermitian, so write its eigendecomposition $\mathbf{A} = \mathbf{W}^\dagger \boldsymbol{\Sigma} \mathbf{W}$. Define $M'_{j,z} \triangleq \mathbf{W} M_{j,z} \mathbf{W}^\dagger$ so that $\sum_{z\in\Omega_j} M'_{j,z} = \mathbb{1}_{d_j}$ and

$$\begin{aligned}
\mathop{\mathbb{E}}_{z\sim q^j}[(g^{\mathbf{U}}_{\mathcal{P}_j}(z) - g^{\mathbf{V}}_{\mathcal{P}_j}(z))^2] &= \mathop{\mathbb{E}}_{z\sim q_j}\left[\left(\frac{1}{\langle M_{j,z}, \sigma_j\rangle}\sum_{i=1}^{d_j}(M'_{j,z})_{ii}\boldsymbol{\Sigma}_{ii}\right)^2\right] \\
&\le \mathop{\mathbb{E}}_{z\sim q_j}\left[\left(\frac{1}{\langle M_{j,z}, \sigma_j\rangle}\sum_{i=1}^{d_j}(M'_{j,z})_{ii}\Sigma^2_{ii}\right)\left(\frac{1}{\langle M_{j,z}, \sigma_j\rangle}\sum_{i=1}^{d_j}(M'_{j,z})_{ii}\right)\right] \\
&\le \frac{1}{p_j}\sum_{z\in\Omega_j}\frac{\mathrm{Tr}(M_{j,z})}{\langle M_{j,z}, \sigma_j\rangle} \cdot \sum_{i=1}^{d_j}(M'_{j,z})_{ii}\boldsymbol{\Sigma}^2_{ii} \\
&\le \frac{1}{p_j}2^{j+1} \cdot \sum_{i=1}^{d_j}\boldsymbol{\Sigma}^2_{ii}\sum_{z\in\Omega_j}(M'_{j,z})_{ii} = \frac{1}{p_j}2^{j+1}\|\boldsymbol{\Sigma}\|^2_{\mathrm{HS}}
\end{aligned}$$

where in the second step we used Cauchy-Schwarz, in the third step we used that $\mathrm{Tr}(M'_{j,z}) = \mathrm{Tr}(M_{j,z})$, in the fourth step we used the fact that the entries of diagonal matrix σ_j are lower bounded by 2^{-j-1}, and in the fifth step we used that $\sum_z \Omega'_{j,z} = \mathbb{1}_{d_j}$. To upper bound $\|\boldsymbol{\Sigma}\|_{\mathrm{HS}}$, note

$$\|\boldsymbol{\Sigma}\|_{\mathrm{HS}} = \|\mathbf{U}^\dagger\boldsymbol{\mathcal{E}}_j\mathbf{U} - \mathbf{U}'^\dagger\boldsymbol{\mathcal{E}}_j\mathbf{U}'\|_{\mathrm{HS}} = \|\mathbf{U}^\dagger\boldsymbol{\mathcal{E}}_j(\mathbf{U}-\mathbf{U}') + (\mathbf{U}'-\mathbf{U})^\dagger\boldsymbol{\mathcal{E}}_j\mathbf{U}'\|_{\mathrm{HS}} \le \varepsilon_j\|\mathbf{U}-\mathbf{U}'\|_{\mathrm{HS}},$$

from which we conclude that $\mathbb{E}_{z\sim q^j}[(g^{\mathbf{U}}_{\mathcal{P}_j}(z) - g^{\mathbf{V}}_{\mathcal{P}_j}(z))^2]^{1/2} \le (2^{j+1}/p_j)^{1/2}\varepsilon_j\|\mathbf{U}-\mathbf{U}'\|_{\mathrm{HS}}$. □

By applying (5) in Lemma 4.4, we get the following bound:

Lemma 5.11. *For any odd t, $\mathbb{E}_{\mathbf{U},\mathbf{V}\sim\mathcal{D}_j}\left[\left(\phi^{\mathbf{U},\mathbf{V}}_{\mathcal{P}_j}\right)^t\right] = 0$, and for any even t,*

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}\sim\mathcal{D}_j}\left[\left(\phi^{\mathbf{U},\mathbf{V}}_{\mathcal{P}_j}\right)^t\right]^{1/t} \le O\left(2^{2j}\varepsilon_j^2/d_j \cdot \left\{\sqrt{t/d_j} \vee t/d_j\right\}\right) \le O\left(t \cdot 2^{2j} \cdot \varepsilon_j^2/d_j^{3/2}\right).$$

Proof. By Lemma 5.8 and the definition of $\phi^{\mathbf{U},\mathbf{V}}_{\mathcal{P}_j}$, $\mathbb{E}[\phi^{\mathbf{U},\mathbf{V}}_{\mathcal{P}_j}] = 0$. By Lemmas 5.9 and 5.10, we can take $\varsigma = O(2^j\varepsilon_j/\sqrt{d_j})$ and $L = O((2^j/p_j)^{1/2}\varepsilon_j)$ when invoking (5) in Lemma 4.4. Note that $p_j \ge d_j 2^{-j-1}$, so $L \le O(\varsigma)$. The claim follows. □

Lemma 5.11, Lemma 3.10, and (9) immediately imply Lemma 5.7.

Proof of Lemma 5.7. From Lemma 3.10, Lemma 5.11, and (9), we have that

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}\sim\mathcal{D}}\left[\left(\phi^{\mathbf{U},\mathbf{V}}\right)^t\right]^{1/t} \le t\left(\sum_{j\in\mathcal{J}} p_j^2 \cdot O\left(\frac{2^{4j}\varepsilon_j^4}{d_j^3}\right)\right)^{1/2} \le t\left(\sum_{j\in\mathcal{J}} O\left(\frac{2^{2j}\varepsilon_j^4}{d_j}\right)\right)^{1/2}$$

where in the second step we used that $p_j \le d_j 2^{-j}$. We can thus expand

$$\mathbb{E}\left[\left(1+\phi^{\mathbf{U},\mathbf{V}}\right)^N\right] = \sum_{2\le t\le N \text{ even}} \binom{N}{t} \mathbb{E}[(\phi^{\mathbf{U},\mathbf{V}})^t] \le \left(\frac{e\cdot N}{t}\right)^t \cdot O\left(t^2 \sum_{j\in\mathcal{J}} \frac{2^{2j}\varepsilon_j^4}{d_j}\right)^{t/2},$$

from which the claim follows by Lemma 4.5. □

5.2.2 Tuning the Perturbations

Before we explain how to tune $\boldsymbol{\mathcal{E}}_j$, we address a minor corner case. Recall from Definition 5.4 that $\boldsymbol{\mathcal{E}}_j$ is zero for buckets j for which $|S_j| = 1$. In the extreme case where all buckets after removal of S_{tail} are of this type, then $\boldsymbol{\mathcal{E}} = 0$ and the problem of distinguishing between σ and $\sigma + \mathbf{U}^\dagger\boldsymbol{\mathcal{E}}\mathbf{U}$ would be vacuous. Fortunately, we can show that if the Schatten 2/5-quasinorm of σ' is dominated by such buckets, then the resulting state certification problem requires many copies because of existing *classical* lower bounds.

Lemma 5.12. *If $\sum_{i\in S_{\mathsf{sing}}\setminus S_{\mathsf{tail}}} \lambda_i^{2/5} \ge \frac{1}{2}\|\sigma'\|_{2/5}^{2/5}$, then state certification with respect to σ using non-adaptive, unentangled measurements has copy complexity at least $\Omega(\|\sigma'\|_{2/5}/\varepsilon^2)$.*

Proof. Intuitively in this case, the spectrum of σ is dominated by eigenvalues in geometric progression, and in fact the instance-optimal lower bound for *classical* identity testing [VV17] already implies a good enough copy complexity lower bound (even against entangled measurements).

Formally, Corollary 3.17 implies a copy complexity lower bound of $\Omega(1/\varepsilon \vee \|\sigma_{-\varepsilon}^{-\max}\|_{2/3}/\varepsilon^2)$. We would like to relate this to

$$\left(\sum_{i\in S_{\mathsf{sing}}\setminus S_{\mathsf{tail}}} \lambda_i^{2/3}\right)^{3/2} \ge (1-2^{-2/5})^{5/2} \cdot \left(\sum_{i\in S_{\mathsf{sing}}\setminus S_{\mathsf{tail}}} \lambda_i^{2/5}\right)^{5/2} \ge \Omega(\|\sigma'\|_{2/5}), \tag{11}$$

where the first step follows by Fact 3.19, and the last step follows by the hypothesis of the lemma.

Suppose that there is some i for which $d_{j(i)} = 1$ and i is not among the indices removed in the definition of $\sigma_{-\varepsilon}^{-\max}$. Then we can lower bound $\|\sigma_{-\varepsilon}^{-\max}\|_{2/3}$ by λ_i, which is at least $(1-2^{-2/3})^{3/2} = \Omega(1)$ times the left-hand side of (11).

On the other hand, suppose that all i for which $d_{j(i)} = 1$ are removed in the definition of $\sigma_{-\varepsilon}^{-\max}$. As long as $\sigma_{-\varepsilon}^{-\max}$ has some nonzero entry, call it λ_{i^*}, then $\lambda_{i^*} \geq \max_{i \in S_{\mathsf{sing}} \backslash S_{\mathsf{tail}}} \lambda_i$, so we can similarly guarantee that $\|\sigma_{-\varepsilon}^{-\max}\|_{2/3} \geq \lambda_{i^*}$ is at least $(1-2^{-2/3})^{3/2} = \Omega(1)$ times the left-hand side of (11). Otherwise, we note that σ' is zero as well, in which case we are also done. □

It remains to consider the primary case where the hypothesis of Lemma 5.12 does not hold, and this is where we will use Lemma 5.7. The following together with Lemma 5.12 will complete the proof of Lemma 5.5:

Lemma 5.13. *If $\sum_{i \in S_{\mathsf{sing}} \backslash S_{\mathsf{tail}}} \lambda_i^{2/5} < \frac{1}{2}\|\sigma'\|_{2/5}^{2/5}$, then state certification with respect to σ using non-adaptive, unentangled measurements has copy complexity at least $\Omega(\|\sigma'\|_{2/5}/(\varepsilon^2 \log(d/\varepsilon)))$.*

The proof of Lemma 5.13 requires some setup. First, obviously the hypothesis of the lemma can equivalently be stated as

$$\sum_{i \in S_{\mathsf{many}} \backslash S_{\mathsf{tail}}} \lambda_i^{2/5} > \frac{1}{2}\|\sigma'\|_{2/5}^{2/5}. \tag{12}$$

Definition 5.14 (Choice of ε_j)**.** *For every $i \in S_{\mathsf{many}}$, for $j \in \mathcal{J}$ the index of the bucket containing i, define $\varepsilon_j \triangleq 2^{-j-1} \wedge \zeta 2^{-2/3(j+1)} d_j^{2/3}$ for normalizing quantity ζ satisfying*

$$\sum_{j \in \mathcal{J} : d_j > 1} 2\lfloor d_j/2 \rfloor \cdot \left\{ 2^{-j-1} \wedge \zeta 2^{-2/3(j+1)} d_j^{2/3} \right\} = \varepsilon. \tag{13}$$

Note that by ensuring that $\varepsilon_j \leq 2^{-j-1}$, we ensure that $\sigma + \mathbf{U}^\dagger \boldsymbol{\mathcal{E}} \mathbf{U}$ has nonnegative spectrum, while (13) ζ ensures that for any $\mathbf{U}$ in the support of $\mathcal{D}$, $\|\boldsymbol{\mathcal{E}}\|_1 = \varepsilon$.

The rest of the proof is devoted to showing that for this choice of $\{\varepsilon_j\}$, the lower bound in (10) is at least the one in Lemma 5.13. The main step is to upper bound the normalizing quantity ζ.

Lemma 5.15. *For ζ defined in Definition 5.14,*

$$\zeta \leq O(\varepsilon) \cdot \left(\sum_{j \in \mathcal{J}', i \in S_j} \lambda_i^{2/3} d_j^{5/3} \right)^{-1}. \tag{14}$$

We will need the following elementary fact (see Appendix B.3 for a proof).

Fact 5.16. *Let $u_1 < \cdots < u_m$ and $v_1 \leq \cdots \leq v_n$ be numbers for which $u_{i+1} \geq 2u_i$ for all i. Let $d_1, \ldots, d_n > 1$ be arbitrary integers. Let $w_1 \leq \cdots \leq w_{m+n}$ be these numbers in sorted order. For $i \in [m+n]$, define d_i^* to be 1 if w_i corresponds to some u_j, and d_j if w_i corresponds to some v_j.*

Let s be the largest index for which $\sum_{i=1}^{s} w_i d_i^* \leq 3\varepsilon$. Let a, b be the largest indices for which u_a, v_b are present among $w_1, \ldots, w_s$ (if none exists, take it to be 0). Then either $b = n$ or $\sum_{i=1}^{b+1} v_i d_i > \varepsilon$.

This allows us to deduce the following bound for buckets not removed in Definition 5.2.

Corollary 5.17. *Under the hypothesis of Lemma 5.13, $S_{\mathsf{many}} \backslash S_{\mathsf{tail}}$ is nonempty, and there exists an absolute constant $c > 0$ such that for any $i \in S_{\mathsf{many}} \backslash S_{\mathsf{tail}}$ in some bucket j, $\zeta \cdot 2^{-2/3(j+1)} d_j^{2/3} \leq c \cdot 2^{-j-1}$.*

Proof. The first part immediately follows from (12). For the second part, take some constant c to be optimized later and suppose to the contrary that for some $i^* \in S_{\mathsf{many}} \backslash S_{\mathsf{tail}}$, lying in some bucket j^*, we have that $c \cdot 2^{-j^*-1} < \zeta \cdot 2^{-2/3(j^*+1)} d_j^{2/3}$, or equivalently $2^{-j^*-1}/d_j^2 < \zeta^3/c^3$. Because in the definition of S_{tail}, we sorted by $\lambda_i/d_{j(i)^2}$, for any $i \in S_{\mathsf{tail}}$, and because $\lambda_i \in [2^{-j(i)-1}, 2^{-j(i)}]$, we also have that $2^{-j(i)-1}/d_{j(i)}^2 < \zeta^3/c^3$, or equivalently, $c \cdot 2^{-j(i)-1} < \zeta \cdot 2^{-2/3(j+1)} d_{j(i)}^{2/3}$.

So the sum on the left-hand side of (13) is at least

$$\sum_{j \in \mathcal{J}: j \geq j^*, d_j > 1} 2\lfloor d_j/2 \rfloor \cdot (c \cdot 2^{-j-1}) \geq \sum_{j \in \mathcal{J}: j \geq j^*, d_j > 1} (2d_j/3) \cdot (c \cdot 2^{-j-1}) \geq \sum_{i \in S_{\mathsf{many}}, i \leq i^*} \lambda_i > \varepsilon,$$

where in the first step we used that for $d_j > 1$, $2\lfloor d_j/2 \rfloor \geq 2d_j/3$, in the second step we took $c = 3$ and used that $\lambda_i \leq 2^{-j}$ for $i \in S_j$, and in the third step we used Fact 5.16 applied to the numbers $\{u_i\} \triangleq \{\lambda_i\}_{i \in S_{\mathsf{sing}}}$, $\{v_i\} \triangleq \{\lambda_i/d_{j(i)}^2\}_{i \in S_{\mathsf{many}}}$ and $\{d_i\} \triangleq \{d_{j(i)}^2\}_{i \in S_{\mathsf{many}}}$. This contradicts (13). □

We are finally ready to upper bound the normalizing constant ζ.

Proof of Lemma 5.15. We can now upper bound ζ as follows. We have

$$\begin{aligned}
\varepsilon &\geq \Omega(\zeta) \cdot \sum_{j \in \mathcal{J}'} 2\lfloor d_j/2 \rfloor \cdot 2^{-2/3(j+1)} d_j^{2/3} \\
&\geq \Omega(\zeta) \sum_{j \in \mathcal{J}'} 2^{-2j/3} d_j^{5/3}
\end{aligned}$$

where in the first step we used (13) and Corollary 5.17, and in the second step we again used the fact that for $d_j > 1$, $2\lfloor d_j/2 \rfloor \geq 2d_j/3$. The claimed bound follows. □

We are now ready to complete the proof of Lemma 5.13:

Proof. Substituting our choice of $\{\varepsilon_j\}$ in Definition 5.14 into the lower bound of Lemma 5.7 gives

$$\begin{aligned}
\left(\sum_{j \in \mathcal{J}} 2^{2j} \|\boldsymbol{\mathcal{E}}_j\|_{\mathrm{op}}^4 / d_j \right)^{-1/2} &\geq \left(\sum_{j \in \mathcal{J}: d_j > 1} \left\{ \frac{2^{-2j-4}}{d_j} \wedge \zeta^4 2^{-2/3j - 8/3} d_j^{5/3} \right\} \right)^{-1/2} \\
&\geq \left(\sum_{j \in \mathcal{J}: d_j > 1} \left\{ \zeta^3 2^{-j-3} d_j \wedge \zeta^4 2^{-2/3j} d_j^{5/3} \right\} \right)^{-1/2} \\
&\geq \Omega(\zeta^{-3/2}) \left(\sum_{j \in \mathcal{J}: d_j > 1} 2\lfloor d_j/2 \rfloor \left\{ 2^{-j-1} \wedge \zeta 2^{-2/3(j+1)} d_j^{2/3} \right\} \right)^{-1/2} \\
&= \Omega(\zeta^{-3/2}) \cdot \varepsilon^{-1/2} \\
&\geq \varepsilon^{-2} \cdot \left(\sum_{j \in \mathcal{J}', i \in S_j} \lambda_i^{2/3} d_j^{5/3} \right)^{3/2} \\
&\geq \max_{j \in \mathcal{J}', i \in S_j} \lambda_i d_j^{5/2} / \varepsilon^2 \\
&\geq \left(\sum_{j \in \mathcal{J}', i \in S_j} \lambda_i^{2/5} d_j \right)^{5/2} \cdot \log(d/\varepsilon)^{-1} \\
&\geq \|\sigma'\|_{2/5} \cdot \log(d/\varepsilon)^{-1},
\end{aligned}$$

where in the second step we used that the minimum of two nonnegative numbers increases if we replace one of them by a weighted geometric mean of the two numbers, in the third step we use the fact that $\lfloor d_j/2 \rfloor$ and d_j are equivalent up to constant factors if $d_j > 1$, in the fourth step we use (13), in the fifth step we use (14), in the penultimate step we used Fact 5.3, and in the last step we used (12) and the fact that for any j, there are at most d_j indices $i \in S_{\mathsf{many}} \backslash S_{\mathsf{tail}}$ within bucket S_j. □

With Lemma 5.13 in place, we conclude the proof of the main lemma of this subsection:

Proof of Lemma 5.5. This follows immediately from Lemmas 5.12 and 5.13. □

5.3 Lower Bound Instance II: Perturbing Off-Diagonals

In many cases, the following lower bound instance will yield a stronger lower bound than the preceding argument, at the cost of applying to a limited range of ε. Take any $j, j' \in \mathcal{J}^*$ for which $d_j \geq d_{j'}$. As we will explain below, if $d_j > 1$, then j and j' need not be distinct.

If j and j' are distinct, then given a matrix $\mathbf{W}^{d_j \times d_{j'}}$ with orthonormal columns, let $\sigma_{\mathbf{W}}$ be the matrix $\sigma + D_{\mathbf{W}}$ where $D_{\mathbf{W}} \in \mathbb{C}^{d \times d}$ is the matrix which is zero outside of the principal submatrix indexed by $S_j \cup S_{j'}$ and which is equal to the matrix

$$\left(\begin{array}{c|c} \mathbf{0}_{d_j} & (\varepsilon/2d_{j'}) \cdot \mathbf{W} \\ \hline (\varepsilon/2d_{j'}) \cdot \mathbf{W}^\dagger & \mathbf{0}_{d_{j'}.} \end{array}\right) \tag{15}$$

On the other hand, if $j = j'$ and $d_j > 1$, then partition S_j into contiguous sets S_j^1, S_j^2 of size $\lceil d_j/2 \rceil$ and $\lfloor d_j/2 \rfloor$, and given a matrix $\mathbf{W}^{\lceil d_j/2 \rceil \times \lfloor d_j/2 \rfloor}$ with orthonormal columns, define $D_{\mathbf{W}} \in \mathbb{C}^{d \times d}$ to be the matrix which is zero outside the principal submatrix indexed by $S_j^1 \times S_j^2$ and which is equal to the matrix

$$\left(\begin{array}{c|c} \mathbf{0}_{\lceil d_j/2 \rceil} & (\varepsilon/2\lfloor d_j/2 \rfloor) \cdot \mathbf{W} \\ \hline (\varepsilon/2\lfloor d_j/2 \rfloor) \cdot \mathbf{W}^\dagger & \mathbf{0}_{d_{\lfloor d_j/2 \rfloor}.} \end{array}\right) \tag{16}$$

In the rest of this subsection, we will consider the case where $j \neq j'$, but as will become evident, all of the following arguments easily extend to the construction for $j = j'$ when $d_j > 1$ by replacing S_j and $S_{j'}$ with S_j^1 and S_j^2 respectively.

Lemma 5.18. *If $\varepsilon \leq d_{j'} \cdot 2^{-j/2-j'/2}$, then $\|\sigma - \sigma_{\mathbf{W}}\|_1 \geq \varepsilon$ and $\sigma_{\mathbf{W}}$ is a density matrix.*

Proof. For the first part, note that

$$\|\sigma - \sigma_{\mathbf{W}}\|_1 = \|D_{\mathbf{W}}\| = 2 \cdot (\varepsilon/2d_{j'})\|\mathbf{W}\|_1 = \varepsilon,$$

where in the second equality we used that $D_{\mathbf{W}}$ is the Hermitian dilation of $(\varepsilon/d_{j'}) \cdot \mathbf{W}$, and in the last equality we used the fact that $\mathbf{W}$ consists of $d_{j'}$ orthogonal columns.

For the second part, first note that regardless of the choice of ε, we have that $\mathrm{Tr}(D_{\mathbf{W}}) = 0$, so $\mathrm{Tr}(\sigma_{\mathbf{W}}) = 1$. Finally, to verify that $\sigma_{\mathbf{W}}$ is positive definite, note that the Schur complement of the principal submatrix of $\sigma_{\mathbf{W}}$ indexed by $S_j \cap S_{j'}$ is given by

$$\sigma_{j'} - \frac{\varepsilon^2}{4d_{j'}^2}\sigma_j^{-1} \succeq 2^{-j'-1}\mathbb{1} - \frac{\varepsilon^2}{4d_{j'}^2}2^{j+1}\mathbb{1},$$

which is positive definite provided that $\varepsilon \leq d_{j'} \cdot 2^{-j/2-j'/2}$. It follows by Lemma 3.14 that $\sigma_{\mathbf{W}}$ is positive definite as claimed. □

The objective of this subsection is to show the following lower bound:

Lemma 5.19. *Fix any $j, j' \in \mathcal{J}^*$ satisfying $d_j \geq d_{j'}$. If $d_j > 1$, then we can optionally take $j = j'$. Suppose $\varepsilon \leq d_{j'} \cdot 2^{-j/2-j'/2}$. Let $\sigma \in \mathbb{C}^{d\times d}$ be a diagonal density matrix. Distinguishing between whether $\rho = \sigma$ or $\rho = \sigma_{\mathbf{W}}$ for $\mathbf{W} \in \mathbb{C}^{d_j \times d_{j'}}$ consisting of Haar-random orthonormal columns, using nonadaptive unentangled measurements, has copy complexity at least*

$$\Omega\left(\frac{\sqrt{d_j} \cdot d_{j'}^2 \cdot 2^{-j'}}{\varepsilon^2}\right).$$

Note that a random $\mathbf{W}$ is equivalent to $\mathbf{U}\Pi$ for $\mathbf{U} \sim \mathcal{D}$, where $\mathcal{D}$ is the Haar measure over $U(d_j)$, and

$$\Pi \triangleq (\mathbb{1}_{d_{j'}} | \mathbf{0}_{d_j - d_{j'}})^\top,$$

so we can just as well parametrize $\{\sigma_{\mathbf{W}}\}$ as $\{\sigma_{\mathbf{U}}\}$, which we will do in the sequel.

Take any single-copy sub-problem $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ where POVM $\mathcal{M}$ consists of elements $\{M_z\}$. Analogously to Lemma 5.6, we may without loss of generality assume that one of the POVM elements is the projector to the coordinates outside of $S_j \cup S_{j'}$, and the remaining POVM elements are rank-1 matrices $M_z = \lambda_z v_z v_z^\dagger$ where the $\lambda_z \leq 1$ satisfy

$$\sum \lambda_z = d_j + d_{j'} < 2d_j \tag{17}$$

and the vectors v_z are unit vectors supported on $S_j \cap S_{j'}$. Let v_z^j and $v_z^{j'}$ denote the d_j- and $d_{j'}$-dimensional components of v_z indexed by S_j and $S_{j'}$. Note that for these z,

$$g_{\mathcal{P}}^{\mathbf{U}}(z) = \frac{\langle M_z, D_{\mathbf{W}}\rangle}{\langle M_z, \sigma\rangle} = \frac{\varepsilon}{d_{j'}} \cdot \frac{\mathrm{Re}((v_z^j)^\dagger(\mathbf{U}\Pi)v_z^{j'})}{v_z^\dagger \sigma v_z}. \tag{18}$$

while for the index z corresponding to the projector to $(S_j \cup S_{j'})^c$, $g_{\mathcal{P}}^{\mathbf{U}}(z) = 0$.

In the next three lemmas, we verify that $\mathcal{P}$ satisfies Assumption 1.

Lemma 5.20. *For any z, $\mathbb{E}_{\mathbf{U}}[g_{\mathcal{P}}^{\mathbf{U}}(z)] = 0$, so Condition 1 of Assumption 1 holds.*

Proof. Clearly $\mathrm{Tr}(D_{\mathbf{W}}) = 0$, so by the second part of Lemma 3.13, $\mathbb{E}_{\mathbf{W}}[g^{\mathbf{W}}(z)] = 0$. □

Lemma 5.21. $\mathbb{E}_{z,\mathbf{U}}[g_{\mathcal{P}}^{\mathbf{U}}(z)^2] \leq O\left(\frac{\varepsilon^2}{d_{j'}^2 2^{-j'}}\right)$*, where as usual, expectation is with respect to measurement outcomes when measuring the null hypothesis σ with $\mathcal{M}$, so Condition 2 of Assumption 1 holds.*

Proof. From (18) we have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{z,\mathbf{U}}\left[g_{\mathcal{P}}^{\mathbf{U}}(z)^2\right] &= \frac{\varepsilon^2}{d_{j'}^2} \mathop{\mathbb{E}}_{\mathbf{U}}\left[\sum_z \lambda_z v_z^\dagger \sigma v_z \left(\frac{\mathrm{Re}((v_z^j)^\dagger(\mathbf{U}\Pi)v_z^{j'})}{v_z^\dagger \sigma v_z}\right)^2\right] \\
&= \frac{\varepsilon^2}{d_{j'}^2} \sum_z \frac{\lambda_z}{v_z^\dagger \sigma v_z} \mathop{\mathbb{E}}_{\mathbf{U}}\left[\left(\mathrm{Re}((v_z^j)^\dagger(\mathbf{U}\Pi)v_z^{j'})\right)^2\right] \\
&= \frac{\varepsilon^2}{d_{j'}^2} \sum_z \frac{\lambda_z}{v_z^\dagger \sigma v_z} \cdot \frac{\|v_z^j\|^2 \|v_z^{j'}\|^2}{d_j},
\end{aligned} \tag{19}$$

As v_z is supported on $S_j \cup S_{j'}$, the supports of v_z^j and $v_z^{j'}$ are disjoint, and the diagonal entries of σ indexed by $S_{j'}$ are at least 2^{-j-1}, we have that $v_z^\dagger \sigma v_z \geq 2^{-j'-1}\|v_z^{j'}\|^2$ and $\|v_z^j\|_2^2 \leq 1$, so we can further bound (19) by

$$= \frac{\varepsilon^2 2^{j'+1}}{d_{j'}^2 d_j} \sum_z \lambda_z \leq O\left(\frac{\varepsilon^2}{d_{j'}^2 2^{-j'}}\right),$$

where the last step follows by (17). □

Lemma 5.22. $\mathbb{E}_z[(g_{\mathcal{P}}^{\mathbf{U}_1}(z) - g_{\mathcal{P}}^{\mathbf{U}_2}(z))^2] \leq O\left(\frac{\varepsilon^2}{d_{j'}^2 2^{-j}}\right) \cdot \|\mathbf{U}_1 - \mathbf{U}_2\|_{HS}^2$ *for any* $\mathbf{U}_1, \mathbf{U}_2 \in U(d_j)$*, so Condition 3 of Assumption 1 holds.*

Proof. Define the matrix

$$\mathbf{D} = \begin{pmatrix} \mathbf{0}_{d_j} & (\varepsilon/2d_{j'}) \cdot (\mathbf{U}_1\Pi - \mathbf{U}_2\Pi) \\ (\varepsilon/2d_{j'}) \cdot (\mathbf{U}_1\Pi - \mathbf{U}_2\Pi)^\dagger & \mathbf{0}_{d_{j'}} \end{pmatrix}$$

Note that for any POVM element M_z,

$$\langle M_z, \mathbf{D}\rangle^2 = \frac{\lambda_z^2\varepsilon^2}{d_{j'}^2} \operatorname{Re}\left((v_z^j)^\dagger(\mathbf{U}_1 - \mathbf{U}_2)\Pi v_z^{j'}\right)^2 \leq \frac{\lambda_z^2\varepsilon^2}{d_{j'}^2} \cdot \|v_z^j(\mathbf{U}_1 - \mathbf{U}_2)\|^2 \cdot \|v_z^{j'}\|_2^2 \tag{20}$$

We can then write

$$\begin{aligned}
\mathop{\mathbb{E}}_z[(g_{\mathcal{P}}^{\mathbf{U}}(z) - g_{\mathcal{P}}^{\mathbf{V}}(z))^2] &= \sum_z \frac{\langle M_z, \mathbf{D}\rangle^2}{\langle M_z, \sigma\rangle} \\
&\leq \frac{\varepsilon^2}{d_{j'}^2} \sum_z \frac{\lambda_z \|v_z^j(\mathbf{U}_1 - \mathbf{U}_2)\|^2 \cdot \|v_z^{j'}\|^2}{2^{-j'-1}\|v_z^{j'}\|^2} \\
&\leq O\left(\frac{\varepsilon^2 2^{j'}}{d_{j'}^2}\right) \cdot \sum_z \lambda_z \|v_z^j(\mathbf{U}_1 - \mathbf{U}_2)\|^2 \\
&= O\left(\frac{\varepsilon^2 2^{j'}}{d_{j'}^2}\right) \cdot \left\langle (\mathbf{U}_1 - \mathbf{U}_2)(\mathbf{U}_1 - \mathbf{U}_2)^\dagger, \sum_z \lambda_z v_z^j (v_z^j)^\dagger \right\rangle \\
&= O\left(\frac{\varepsilon^2}{d_{j'}^2 2^{-j'}}\right) \cdot \|\mathbf{U}_1 - \mathbf{U}_2\|_{\mathrm{HS}}^2,
\end{aligned}$$

where in the second step we used (20) and the fact that $\langle M_z, \sigma\rangle = \lambda_z v_z^\dagger \sigma v_z \geq \lambda_z 2^{-j'-1}\|v_z^{j'}\|^2$, and in the fifth step we used that $\sum_z \lambda_z v_z^j (v_z^j)^\dagger = \mathbb{1}_{d_j}$. □

We can finally complete the proof of Lemma 5.19:

Proof of Lemma 5.19. As the mixture of alternatives in $\mathcal{P}$ is parametrized by $\mathbf{U} \sim \mathcal{D}$ for $\mathcal{D}$ the Haar measure over the unitary group, the lemma immediately follows from Lemma 4.6 with $L, \varsigma = O\left(\frac{\varepsilon}{d_{j'} 2^{-j'/2}}\right)$. □

5.4 Lower Bound Instance III: Corner Case

We will also need the a lower bound instance that will yield an $\Omega(1/\varepsilon^2)$ lower bound for state certification with respect to any σ with maximum entry at least $1/2$. We will not use anything about bucketing in this warmup result.

Let i_1 be the index of the largest entry of σ, and let i_2 be the index of the second-largest (breaking ties arbitrarily). For any $u \in \{\pm 1\}$, consider the state σ^u which agrees with σ everywhere except in the principal submatrix indexed by $\{i_1, i_2\}$. Within that submatrix, define $\sigma^u_{i_1,i_1} = \sigma_i - \varepsilon^2/4$, $\sigma^u_{i_2,i_2} = \sigma_{i_2} + \varepsilon^2/4$, and $\sigma^u_{i_1,i_2} = {\sigma^u}^\dagger_{i_2,i_1} = (\varepsilon/2)u$.

Lemma 5.23. *If the maximum entry of σ is at least $3/4$, then for any $\varepsilon \le 1/2$, $\|\sigma - \sigma^u\|_1 \ge \varepsilon$ and σ^u is a density matrix.*

Proof. Note that for $\varepsilon < 1/2$,

$$\|\sigma - \sigma^u\|_1 = \left\| \begin{pmatrix} -\varepsilon^2 & (\varepsilon/2)u \\ (\varepsilon/2)\overline{u} & \varepsilon^2 \end{pmatrix} \right\|_1 = 2\sqrt{\varepsilon^4/16 + \varepsilon^2/4} \ge \varepsilon.$$

For the second part of the lemma, clearly $\mathrm{Tr}(\sigma^u) = 1$. To verify that σ^u is psd, first note that because $\sigma_{i_1,i_1} \ge 3/4$ and $\sigma_{i_2,i_2} \le 1/2$, and $\varepsilon^2/4 \le 1/4$, every diagonal entry of σ^u is nonnegative. On the other hand, the principal submatrix indexed by $\{i_1, i_2\}$ has determinant $(\sigma_{i_1,i_1} - \varepsilon^2)(\sigma_{i_2,i_2} + \varepsilon^2) - \varepsilon^2/4 \ge (3/4 - \varepsilon^2)\varepsilon^2 - \varepsilon^2/4 \ge 0$, so σ^u is psd as claimed. □

The objective of this subsection is to show the following lower bound:

Lemma 5.24. *Let $\varepsilon \le 1/2$. If the maximum entry of σ is at least $3/4$, then distinguishing between whether $\rho = \sigma$ or $\rho = \sigma^u$ for $u \sim \{\pm 1\}$, using nonadaptive unentangled measurements, has copy complexity at least $\Omega(1/\varepsilon^2)$. In fact, this holds even for adaptive unentangled measurements.*

Because we have no a priori bound on σ_{i_2,i_2}, the KL divergence between the distribution over outcomes from measuring N copies of σ^u for random $u \in \{\pm 1\}$ and the distribution from measuring N copies of σ may be arbitrarily large, so we cannot implement the strategy in Section 4. Instead, we will directly upper bound the total variation between these two distributions using the following basic fact:

Fact 5.25. *Given distributions p, q over a discrete domain S, if likelihood ratio $p(x)/q(x) \ge 1 - \nu$, then $d_{\mathrm{TV}}(p, q) \le \nu$.*

Proof. We can write

$$d_{\mathrm{TV}}(p,q) = \sum_{x : p(x) \le q(x)} |p(x) - q(x)| = \sum_{x : p(x) \le q(x)} q(x) \cdot |p(x)/q(x) - 1| \le \nu$$

as claimed. □

Proof of Lemma 5.24. Let $\mathcal{D}$ be the uniform distribution over $\{\pm 1\}$, and fix an arbitrary unentangled POVM schedule $\mathcal{S}$. Let p_0 denote the distribution over transcripts $z_{\le t}$ of outcomes upon measuring N copies of σ with $\mathcal{S}$, and let p_1 denote the distribution upon measuring N copies of σ^u, where $u \sim \mathcal{D}$. We will lower bound the likelihood ratio $p_1(z_{\le N})/p_0(z_{\le N})$ for *any* transcript $z_{\le N}$. Let $\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(N)}$ denote the (possibly adaptively chosen) POVMs that were used in the course of generating $z_{\le N}$.

For any $t \in [N]$, suppose $\mathcal{M}^{(t)}$ consists of elements $\{M_z^{(t)}\}$. Analogously to Lemma 5.6, we may without loss of generality assume that one element of $\mathcal{M}^{(t)}$ is the projector to the coordinates outside of $\{i_1, i_2\}$, and the remaining elements are rank-1 matrices $M_z^{(t)} = \lambda_z^{(t)} v_z^{(t)} (v_z^{(t)})^\dagger$ where the $\lambda_z^{(t)} \le 1$ satisfy $\sum \lambda_z^{(t)} = 2$ and the vectors $v_z^{(t)}$ are unit vectors supported on $\{i_1, i_2\}$. Let $v_{z_t,1}^{(t)}$ and $v_{z_t,2}^{(t)}$ denote the coordinates of $v_z^{(t)}$ indexed by i_1 and i_2.

Note that for any $u \in \{\pm 1\}$ and $t \in [N]$, if z_t does not correspond to the projector to the coordinates outside of $\{i_1, i_2\}$, we can write

$$\Delta_t^u(z_t) \triangleq \frac{\langle M_{z_t}^{(t)}, \sigma^u \rangle}{\langle M_{z_t}^{(t)}, \sigma \rangle} = 1 + \frac{\varepsilon u \operatorname{Re}\left(\overline{v_{z_t,1}^{(t)}} v_{z_t,2}^{(t)}\right) - \varepsilon^2 \left(\left|v_{z_t,1}^{(t)}\right|^2 - \left|v_{z_t,2}^{(t)}\right|^2\right)}{v_{z_t}^{(t)\dagger} \sigma v_{z_t}^{(t)}}$$

and if z_t does correspond to the projector, then $\Delta_t^u(z_t) = 1$.

Denoting the t-th entry of $z_{\le N}$ by z_t, we can use AM-GM to bound the likelihood ratio by

$$\begin{aligned} \frac{p_1(z_{\le N})}{p_0(z_{\le N})} &= \mathop{\mathbb{E}}_u \left[\prod_{t=1}^N \Delta_t^u(z_t)\right] \\ &\ge \left(\prod_{t=1}^N \Delta_t^{+1}(z_t)\Delta_t^{-1}(z_t)\right)^{1/2} \end{aligned} \tag{21}$$

To prove the lemma, we will lower bound this by $1 - o(1)$. Because $\Delta_t^u(z_t) = 1$ if z_t corresponds to the projector to the coordinates outside of $\{i_1, i_2\}$, we may assume without loss of generality that this is not the case for any $t \in [N]$. We can then further bound (21) by

$$\ge \prod_{t=1}^N \left\{ \left(1 - \frac{\varepsilon^2 \left(\left|v_{z_t,1}^{(t)}\right|^2 - \left|v_{z_t,2}^{(t)}\right|^2\right)}{v_{z_t}^{(t)\dagger} \sigma v_{z_t}^{(t)}}\right)^2 - \frac{\varepsilon^2 \operatorname{Re}\left(\overline{v_{z_t,1}^{(t)}} v_{z_t,2}^{(t)}\right)^2}{\left(v_{z_t}^{(t)\dagger} \sigma v_{z_t}^{(t)}\right)^2} \right\}^{1/2}. \tag{22}$$

For any $v \in \mathbb{C}^d$ which has entries v_1 and v_2 in coordinates i_1 and i_2 and is zero elsewhere, we have that

$$\frac{|v_1|^2 - |v_2|^2}{v^\dagger \sigma v} \le \frac{|v_1|^2}{\sigma_{i_1,i_1}|v_1|^2} \le 4/3 \qquad \frac{\operatorname{Re}(\overline{v_1} v_2)^2}{v^\dagger \sigma v} \le \frac{\operatorname{Re}(\overline{v_1} v_2)^2}{\sigma_{i_1,i_1}|v_1|^2} \le 4/3,$$

where the last step for both estimates follows by the assumed lower bound on σ_{i_1,i_1}. By (22) we have that

$$\frac{p_1(z_{\le N})}{p_0(z_{\le N})} \ge ((1 - 4\varepsilon^2/3)^2 - 4\varepsilon^2/3)^{N/2} \ge (1 - 32\varepsilon^2/9)^{N/2}.$$

In particular, for $N = o(1/\varepsilon^2)$, the likelihood ratio is at least $1 - o(1)$ as desired. □

5.5 Putting Everything Together

We are now ready to conclude the proof of Theorem 5.1.

Proof of Theorem 5.1. We proceed by casework depending on whether or not $d_j = 1$ for all $j \in \mathcal{J}^*$.

Case 1. *$d_j = 1$ for all $j \in \mathcal{J}^*$.*

There are two possibilities. If there is a single bucket $j = j(i)$ for which $i \notin S_{\mathsf{tail}} \cup S_{\mathsf{light}}$, then $d_{\mathsf{eff}} = 1$ and $\|\sigma^{**}\|_{1/2} = O(1)$. For ε smaller than some absolute constant, we know that $\sigma_{i,i} \geq 3/4$ and can apply Lemma 5.24 to conclude a lower bound of $\Omega(1/\varepsilon^2)$ as desired. Otherwise, let j' be the smallest index for which $j' = j(i')$ for some $i' \in \mathcal{J}^*$, and let $j > j'$ be the next smallest index for which $j = j(i)$ for some $i \in \mathcal{J}^*$. Consider the lower bound instance in Section 5.3 applied to this choice of j, j'. Provided that $\varepsilon \leq 2^{-j/2-j'/2}$, we would obtain a copy complexity lower bound of $\Omega(2^{-j'}/\varepsilon^2) \geq \Omega(\|\sigma^*\|_{1/2}/(\varepsilon^2 \log(d/\varepsilon)))$, where the inequality is by Fact 3.18, and we would be done. On the other hand, if $\varepsilon \geq 2^{-j/2-j'/2}$, then because $2^{-j'} > 2^{-j}$, we would conclude that $2^{-j} \leq \varepsilon$. In particular, this implies that $\sum_{j'' \in \mathcal{J}^*, i \in S_{j''} : j'' \neq j'} \lambda_i \leq 2\varepsilon$, so after removing at most an additional 2ε mass from σ^*, we get a matrix σ^{**} (see Definition 5.2) with a single nonzero entry. Again, $d_{\mathsf{eff}} = 1$ and $\|\sigma^{**}\|_{1/2} = O(1)$, and if ε is smaller than some absolute constant, we conclude that that single nonzero entry is at least $3/4$ and can apply Lemma 5.24 to conclude a lower bound of $\Omega(1/\varepsilon^2)$ as desired.

Case 2. *$d_j > 1$ for some $j \in \mathcal{J}^*$.*

Let $j_* \triangleq \arg\max_{j \in \mathcal{J}^*} d_j$ and $j'_* \triangleq \arg\max_{j \in \mathcal{J}^*} d_j^2 2^{-j}$. By Lemma 5.19, we have a lower bound of $\Omega\left(\sqrt{d_{j_*}} \cdot d_{j'_*}^2 \cdot 2^{-j'_*}/\varepsilon^2\right)$ as long as ε satisfies the bound

$$\varepsilon \leq d_{j'_*} \cdot 2^{-j_*/2 - j'_*/2}. \tag{23}$$

Note that because $d_{j_*} > 1$ as we are in Case 2, we do not constrain j_*, j'_* to be distinct necessarily. We would now like to argue that this lower bound, up to log factors, holds even if the bound on ε in (23) does not hold. In the following, assume that (23) does not hold.

To this end, we will also use the lower bound from Lemma 5.5 of $\Omega(\|\sigma'\|_{2/5}/(\varepsilon^2 \log(d/\varepsilon)))$. We would first like to relate $\|\sigma'\|_{2/5}$ to $\|\sigma^*\|_{2/5}$.

Lemma 5.26. *Either $\|\sigma'\|_{2/5} \geq \Omega(\|\sigma^*\|_{2/5})$, or the following holds. Let j° be the index maximizing $d_j^{5/2} 2^{-j}$. Then 1) $j^\circ = \min_{j \in \mathcal{J}^*} j$, 2) $d_{j^\circ} = 1$, and 3) $j^\circ = 0$.*

Proof. We will assume that $\|\sigma'\|_{2/5} = o(\|\sigma^*\|_{2/5})$ and show that 1), 2), and 3) must hold. Let j° be the index maximizing $d_j^{5/2} 2^{-j}$, and let $i_{\max}$ be the index of the top entry of σ^*. Let σ'' denote the matrix obtained by zeroing out the top entry of σ^*. Note that the nonzero entries of σ' comprise a superset of those of σ'', so

$$\frac{\|\sigma^*\|_{2/5}^{2/5}}{\|\sigma'\|_{2/5}^{2/5}} \leq \frac{\|\sigma^*\|_{2/5}^{2/5}}{\|\sigma''\|_{2/5}^{2/5}} = \frac{\sum_{i \in \mathcal{J}^*} \sigma_i^{2/5}}{\sum_{i \in \mathcal{J}^* \setminus \{i_{\max}\}} \sigma_i^{2/5}}.$$

Suppose 1) does not hold. Then

$$\frac{\sum_{i \in \mathcal{J}^*} \sigma_i^{2/5}}{\sum_{i \in \mathcal{J}^* \setminus \{i_{\max}\}} \sigma_i^{2/5}} \leq \frac{\sigma_{i_{\max}}^{2/5} + \sum_{i \in S_{j^\circ}} \sigma_i^{2/5}}{\sum_{i \in S_{j^\circ}} \sigma_i^{2/5}} \leq 2,$$

where the first inequality follows by the elementary fact that for positive integers $a \geq b$ and c, $\frac{a+c}{b+c} \leq \frac{a}{b}$, and the second inequality follows by the definition of j°.

Next, suppose 1) holds but 2) does not hold. Then

$$\frac{\sum_{i \in \mathcal{J}^*} \lambda_i^{2/5}}{\sum_{i \in \mathcal{J}^* \setminus \{i_{\max}\}} \lambda_i^{2/5}} \leq \frac{\sum_{i \in S_{j^\circ}} \lambda_i^{2/5}}{\sum_{i \in S_{j^\circ} \setminus \{i_{\max}\}} \lambda_i^{2/5}} \leq O(1),$$

where the first inequality again uses the above elementary fact, the second inequality follows by our assumption that 2) does not hold. This yields a contradiction.

Finally suppose 1) and 2) hold, but 3) does not, so that $\|\sigma^*\|_\infty \le 1/2$. Let σ'' denote the matrix obtained by zeroing out the top entry of σ^*. We would have

$$\|\sigma''\|_{2/5} \ge \|\sigma''\| \ge 1/2 - O(\varepsilon),$$

so for ε smaller than a sufficiently large absolute constant, we would have that $\|\sigma''\|_{2/5}^{2/5} \ge \Omega(\|\sigma^*\|_\infty^{2/5})$ and therefore $\|\sigma'\|_{2/5} \ge \|\sigma''\|_{2/5} \ge \Omega(\|\sigma^*\|_{2/5})$, a contradiction. □

Suppose the latter scenario in Lemma 5.26 happens but the former does not. In this case, because $d_{j^\circ} = 1$, we also have that $j'_* = \arg\max_{j \in \mathcal{J}^*} d_j^2 2^{-j}$, i.e. $j'_* = j^\circ$. In particular,

$$1 \ge d_{j'_*}^2 2^{-j'_*} \ge d_{j_*}^2 2^{-j_*} \ge \Omega(d_{j_*}^{3/2} \varepsilon / \log(d/\varepsilon)), \tag{24}$$

where the last inequality follows by the fact that $d_j 2^{-j} \ge \Omega(\varepsilon/\log(d/\varepsilon))$ for all $j \in \mathcal{J}^*$ by design. We conclude that $\varepsilon \le O(d_{j_*}^{-3/2} \log(d/\varepsilon))$. But recall that we are assuming that (23) is violated, i.e. that

$$\varepsilon > d_{j'_*} \cdot 2^{-j_*/2 - j'_*/2} = 2^{-j_*/2 - j'_*/2} \ge \Omega(\varepsilon/(d_{j_*} \log(d/\varepsilon)))^{1/2}, \tag{25}$$

where the last step is by 3) in Lemma 5.26 and the fact that $d_j 2^{-j} \ge \Omega(\varepsilon/\log(d/\varepsilon))$ for all $j \in \mathcal{J}^*$. Combining (24) and (25), we get a contradiction of the assumption that the former scenario in Lemma 5.26 does not hold, unless $d_{j_*} \le \mathrm{polylog}(d/\varepsilon)$. But if $d_{j_*} \le \mathrm{polylog}(d/\varepsilon)$, then the lower bound claimed in Theorem 5.1 still holds as $d_{\mathsf{eff}} \le O(\log(d/\varepsilon) \cdot d_{j_*}) \le \mathrm{polylog}(d/\varepsilon)$.

Finally, suppose instead that the former scenario in Lemma 5.26 happens, so that Lemma 5.5 gives a lower bound of $\Omega(\|\sigma^*\|_{2/5}/(\varepsilon^2 \log(d/\varepsilon)))$. Let j° still be as defined in Lemma 5.26.

Now we would certainly be done if this lower bound were, up to log factors, larger than the one guaranteed by Lemma 5.19 to begin with. So suppose to the contrary. We would get that

$$d_{j_*}^{5/2} 2^{-j_*} \le d_{j^\circ}^{5/2} 2^{-j^\circ} \le \frac{1}{\log^2(d/\varepsilon)} \sqrt{d_{j_*}} d_{j'_*}^2 \cdot 2^{-j'_*},$$

implying that

$$d_j^2 2^{-j_*} \le \frac{1}{\log^2(d/\varepsilon)} d_{j'_*}^2 2^{-j'_*}. \tag{26}$$

If (23) does not hold, then

$$\frac{1}{\log(d/\varepsilon)} \cdot d_{j'_*} \cdot 2^{-j_*/2 - j'_*/2} \le \frac{\varepsilon}{\log(d/\varepsilon)} \le d_j 2^{-j},$$

where in the last step we again used the fact that $d_j 2^{-j} > \varepsilon/\log(d/\varepsilon)$ for all $j \in \mathcal{J}^*$, yielding the desired contradiction with (26) upon rearranging.

Having lifted the constraint (23), we finally note that by Fact 3.18,

$$\Omega\left(\sqrt{d_{j_*}} \cdot d_{j'_*}^2 \cdot 2^{-j'_*/\varepsilon^2}\right) \ge \Omega\left(\sqrt{d_{\mathsf{eff}}} \cdot \|\sigma^*\|_{1/2}/(\varepsilon^2 \,\mathrm{polylog}(d/\varepsilon))\right).$$

The proof is complete upon invoking Fact 5.27 below. □

Fact 5.27. *Given psd matrix $\sigma \in \mathbb{C}^{d \times d}$, let $\widehat{\sigma} \triangleq \sigma/\operatorname{Tr}(\sigma)$. Then $\|\sigma\|_{1/2} = d \operatorname{Tr}(\sigma)^2 \cdot F(\widehat{\sigma}, \rho_{\mathsf{mm}})$.*

Proof. We may assumed without loss of generality that σ is diagonal. By definition

$$F(\widehat{\sigma}, \rho_{\mathsf{mm}}) = \left(\operatorname{Tr}\sqrt{\sqrt{\widehat{\sigma}}(\mathbb{1}/d)\sqrt{\widehat{\sigma}}}\right)^2 = \left(\frac{1}{\sqrt{d}\operatorname{Tr}(\sigma)} \cdot \operatorname{Tr}(\sqrt{\sigma})\right)^2 = \frac{1}{d\operatorname{Tr}(\sigma)^2} \cdot \|\sigma\|_{1/2},$$

from which the claim follows. □

6 State Certification Algorithm

In this section we prove the following upper bound on state certification that nearly matches the lower bound proven in Section 5:

Theorem 6.1. *Fix $\varepsilon, \delta > 0$. Let $\rho \in \mathbb{C}^{d \times d}$ be an unknown mixed state, and let $\sigma \in \mathbb{C}^{d \times d}$ be a diagonal density matrix. Let σ' be the matrix given by zeroing out the bottom $O(\varepsilon^2)$ mass in σ (see Definition 6.5 below). Let $\widehat{\sigma}' \triangleq \sigma' / \operatorname{Tr}(\sigma')$ and let d_{eff} be the number of nonzero entries of σ'.*

Given an explicit description of σ and copy access to ρ, CERTIFY *takes*

$$N = O(d\sqrt{d_{\mathsf{eff}}} \cdot F(\widehat{\sigma}', \rho_{\mathsf{mm}}) \operatorname{polylog}(d/\varepsilon) \log(1/\delta)/\varepsilon^2)$$

copies of ρ and, using unentangled nonadaptive measurements, distinguishes between $\rho = \sigma$ and $\|\rho - \sigma\|_1 > \varepsilon$ with probability at least $1 - \delta$.

First, in Section 6.1 we give a generic algorithm for state certification based on measuring in a Haar-random basis and applying classical identity testing. In Section 6.2, we describe a bucketing scheme that will be essential to the core of our analysis in Section 6.3, where we use this tool to obtain the algorithm in Theorem 6.1.

6.1 Simple Subroutine

The main result of this section is a basic state certification algorithm that will be invoked as a subroutine in our instance-near-optimal certification algorithm:

Lemma 6.2. *Fix $\varepsilon, \delta > 0$. Let $\rho, \sigma \in \mathbb{C}^{d \times d}$ be two mixed states. Given access to an explicit description of σ and copy access to ρ,* BASICCERTIFY *takes $N = O(\sqrt{d} \log(1/\delta)/\varepsilon^2)$ copies of ρ and, using unentangled nonadaptive measurements, distinguishes between $\rho = \sigma$ and $\|\rho - \sigma\|_{HS} > \varepsilon$ with probability at least $1 - \delta$.*

Algorithm 1: BASICCERTIFY$(\rho, \sigma, \varepsilon, \delta)$

Input: Copy access to ρ, diagonal density matrix σ, error ε, failure probability δ
Output: YES if $\rho = \sigma$, NO if $\|\rho - \sigma\|_{\mathrm{HS}} > \varepsilon$, with probability $1 - \delta$

$N \leftarrow O(\sqrt{d}/\varepsilon^2)$.
for $T = 1, \ldots, O(\log(1/\delta))$ **do**
  Sample a Haar-random unitary matrix $\mathbf{U}$.
  Form the POVM $\mathcal{M}$ consisting of $\{|\mathbf{U}_1\rangle\langle\mathbf{U}_1|, \ldots, |\mathbf{U}_d\rangle\langle\mathbf{U}_d|\}$.
  Measure each copy of ρ with $\mathcal{M}$, yielding outcomes $z_1, \ldots, z_N$.
  Let $q \in \Delta^d$ denote the distribution over outcomes from measuring σ with $\mathcal{M}$.
  Draw i.i.d. samples $z'_1, \ldots, z'_N$ from q.
  $b_i \leftarrow$ L2TESTER$(\{z_i\}, \{z'_i\})$.
return majority among $b_1, \ldots, b_T$.

To prove Lemma 6.2, we will need the following result from classical distribution testing.

Lemma 6.3 (Lemma 2.3 from [DK16])**.** *Let p, q be two unknown distributions on $[d]$ for which $\|p\|_2 \wedge \|q\|_2 \leq b$ for some $b > 0$. There exists an algorithm* L2TESTER *that takes $N = O(b \log(1/\delta)/\varepsilon^2)$ samples from each of p and q and distinguishes between $p = q$ and $\|p - q\|_2 > \varepsilon$ with probability at least $1 - \delta$.*[6]

[6]Note that Lemma 2.3 in [DK16] only gives a constant probability guarantee, but the version we state follows by a standard amplification argument.

We will also need the following moment calculations:

Lemma 6.4. *For any Hermitian $\mathbf{M} \in \mathbb{C}^{d\times d}$ and Haar-random $\mathbf{U} \in U(d)$, let Z denote the random variable $\sum_{i=1}^{d} \left(\mathbf{U}_i^\dagger \mathbf{M} \mathbf{U}_i\right)^2$. Then*

$$\mathbb{E}[Z] = \frac{1}{d+1}\left(\mathrm{Tr}(\mathbf{M})^2 + \|\mathbf{M}\|_{HS}^2\right).$$

If in addition we have that $\mathrm{Tr}(\mathbf{M}) = 0$, then

$$\mathbb{E}[Z^2] \leq \frac{1+o(1)}{d^2}\|\mathbf{M}\|_{HS}^4.$$

Proof. By symmetry $\mathbb{E}[Z] = d\,\mathbb{E}[(\mathbf{U}_1\mathbf{M}\mathbf{U}_1)^2]$, and by Lemma 3.13, if Π denotes the projector to the first coordinate,

$$\mathbb{E}[(\mathbf{U}_1\mathbf{M}\mathbf{U}_1)^2] = \sum_{\pi,\tau\in S_2} \mathrm{Wg}(\pi\tau^{-1}, d)\langle\Pi\rangle_\pi \langle\mathbf{M}\rangle_\tau = \frac{1}{d(d+1)}(\mathrm{Tr}(\mathbf{M})^2 + \mathrm{Tr}(\mathbf{M}^2)),$$

from which the first part of the lemma follows.

For the second part, let $\mathcal{S}_4^* \subset S_4$ denote the set of permutations π for which $\pi(1), \pi(2) \in \{1,2\}$ and $\pi(3), \pi(4) \in \{3,4\}$. Note that

$$\mathbb{E}[Z^2] = d \cdot \mathbb{E}\Big[(\mathbf{U}_1^\dagger \mathbf{M}\mathbf{U}_1)^4\Big] + (d^2-d)\cdot \mathbb{E}\Big[(\mathbf{U}_1^\dagger\mathbf{M}\mathbf{U}_1)^2(\mathbf{U}_2^\dagger\mathbf{M}\mathbf{U}_2)^2\Big]. \tag{27}$$

For the first term, by Lemma 3.13 we have

$$\begin{aligned}
\mathbb{E}[(\mathbf{U}_1^\dagger\mathbf{M}\mathbf{U}_1)^4] &= \sum_{\pi,\tau\in S_4} \mathrm{Wg}(\pi\tau^{-1}, d)\langle\mathbf{M}\rangle_\tau \\
&= \frac{1}{d(d+1)(d+2)(d+3)} \sum_{\tau} \langle\mathbf{M}\rangle_\tau \\
&= \frac{1}{d(d+1)(d+2)(d+3)} \sum_{\tau \text{ derangement}} \langle\mathbf{M}\rangle_\tau \\
&\leq \frac{O(\|\mathbf{M}\|_{\mathrm{HS}}^4)}{d(d+1)(d+2)(d+3)},
\end{aligned}$$

where the third step follows by the fact that $\mathrm{Tr}(\mathbf{M}) = 0$, and the fourth by the fact that for any derangement $\tau \in S_4$, either $\langle\mathbf{M}\rangle_\tau = \mathrm{Tr}(\mathbf{M}^2)^2 = \|\mathbf{M}\|_{\mathrm{HS}}^4$, or $\langle\mathbf{M}\rangle_\tau = \mathrm{Tr}(\mathbf{M}^4) \leq \|\mathbf{M}\|_{\mathrm{HS}}^4$. Similarly,

$$\begin{aligned}
\mathbb{E}[(\mathbf{U}_1^\dagger\mathbf{M}\mathbf{U}_1)^2(\mathbf{U}_2^\dagger\mathbf{M}\mathbf{U}_2)^2] &= \sum_{\pi\in\mathcal{S}_4^*,\tau\in S_4} \mathrm{Wg}(\pi\tau^{-1}, d)\langle\mathbf{M}\rangle_\tau \\
&= \sum_{\tau\in\mathcal{S}_4^*} \mathrm{Wg}(e, d)\langle\mathbf{M}\rangle_\tau + \sum_{\pi\in\mathcal{S}_4^*,\tau\in S_4:\tau\neq\pi} \mathrm{Wg}(\pi\tau^{-1}, d)\langle\mathbf{M}\rangle_\tau \\
&= \mathrm{Wg}(e, d)\|\mathbf{M}\|_{\mathrm{HS}}^4 + \sum_{\pi\in\mathcal{S}_4^*,\tau\in S_4:\tau\neq\pi} \mathrm{Wg}(\pi\tau^{-1}, d)\langle\mathbf{M}\rangle_\tau \\
&\leq \frac{d^4-8d^2+6}{d^2(d^6-14d^4+49d^2-36)}\|\mathbf{M}\|_{\mathrm{HS}}^4 + O(1/d^5)\cdot\|\mathbf{M}\|_{\mathrm{HS}}^4 \\
&= \frac{1+o(1)}{d^4}\|\mathbf{M}\|_{\mathrm{HS}}^4,
\end{aligned}$$

where in the second step $\mathrm{Wg}(e,d)$ denotes the Weingarten function corresponding to the identity permutation, in the third step we used the fact that the only $\tau \in \mathcal{S}_4^*$ which is a derangement is the permutation that interchanges 1 with 2, and 3 with 4, and in the fourth step we used the form of $\mathrm{Wg}(e,d)$, the fact that $|\mathrm{Wg}(\pi\tau^{-1},d)| = O(1/d^5)$ for $\pi \neq \tau$, and the fact that $\langle \mathbf{M} \rangle_\tau \leq \|\mathbf{M}\|_{\mathrm{HS}}^4$. The second part of the lemma follows from (27). □

We can now complete the proof of Lemma 6.2.

Proof of Lemma 6.2. Let p and q be the distribution over d outcomes when measuring ρ and σ respectively using the POVM defined in a single iteration of the main loop of BasicCertify. Applying both parts of Lemma 6.4 to $\mathbf{M} = \rho - \sigma$, for which the random variable Z is $\|p-q\|_2^2$, we conclude that for some sufficiently small absolute constant $c > 0$, $\Pr[\|p-q\|_2 \geq c\|\mathbf{M}\|_{\mathrm{HS}}/\sqrt{d}] \geq 5/6$. Applying the first part of Lemma 6.4 to $\mathbf{M} = \rho$ and $\mathbf{M} = \sigma$, for which the random variable Z is $\|p\|_2^2$ and $\|q\|_2^2$ respectively, we have that $\mathbb{E}[\|p\|_2^2], \mathbb{E}[\|q\|_2^2] \leq 2/d$, so by Markov's, for some absolute constant $c' > 0$, $\|p\|, \|q\|_2 \leq c'/\sqrt{d}$ with probability at least 5/6. We can substitute these bounds for $\|p\|_2, \|q\|_2, \|p-q\|_2$ into Lemma 6.3 to conclude that the output of L2Tester is correct with some constant advantage. Repeating this $O(\log(1/\delta))$ times and taking the majority among all the outputs from L2Tester gives the desired high-probability guarantee. □

6.2 Bucketing and Mass Removal

We may without loss of generality assume that σ is the diagonal matrix $\mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, where $\lambda_1 \leq \cdots \leq \lambda_d$.

We will use the bucketing procedure outlined in Section 5.1. The way that we remove a small amount of mass from the spectrum of σ slightly differs from that outlined in Definition 5.2 for our lower bound. Our bucketing and mass removal procedure is as follows:

Definition 6.5 (Removing low-probability elements- upper bound)**.** *Let $d' \leq d$ denote the largest index for which $\sum_{i=1}^{d'} \lambda_i' \leq \varepsilon^2/20$,[7] and let $S_{\mathsf{tail}} \triangleq [d']$. Let σ' denote the matrix given by zeroing out the diagonal entries of σ indexed by S_{tail}. For $j \in \mathbb{Z}_{\geq 0}$, let S_j denote the indices $i \notin S_{\mathsf{tail}}$ for which $\lambda_i \in [2^{-j-1}, 2^{-j}]$, and denote $|S_j|$ by d_j. Let $\mathcal{J}$ denote the set of j for which $S_j \neq \emptyset$.*

As in the proofs of our lower bounds, we use the following basic consequence of bucketing:

Fact 6.6. *There are at most $\log(10d/\varepsilon^2)$ indices $j \in \mathcal{J}$.*

Proof. The largest element among $\{\lambda_i\}_{i \in S_{\mathsf{tail}}}$ is at least $\varepsilon^2/10d$, from which the claim follows. □

We now introduce some notation. Let $m \triangleq \log(10d/\varepsilon^2)$ denote this upper bound on the number of buckets in $\mathcal{J}$. For $j \in \mathcal{J}$, let $\rho[j,j], \sigma[j,j] \in \mathbb{C}^{d \times d}$ denote the Hermitian matrices given by zeroing out entries of ρ, σ outside of the principal submatrix indexed by S_j. For distinct $j, j' \in \mathcal{J}$, let $\rho[j,j'] \in \mathbb{C}^{d \times d}$ denote the Hermitian matrix given by zeroing out entries of ρ outside of the two non-principal submatrices with rows and columns indexed by S_i and S_j, and by S_j and S_i. Lastly, let $\widehat{\rho}[j,j], \widehat{\sigma}[j,j], \widehat{\rho}[j,j'], \widehat{\sigma}[j,j']$ denote these same matrices but with trace normalized to 1.

Let $\rho_{\mathsf{junk}}^{\mathsf{diag}} \in \mathbb{C}^{d \times d}$ be the principal submatrix of ρ indexed by S_{tail}, and let $\rho_{\mathsf{junk}}^{\mathsf{off}} \in \mathbb{C}^{d \times d}$ be the matrix given by zeroing out the principal submatrices indexed by S_{tail} and by $[d] \backslash S_{\mathsf{tail}}$.

Lastly, we will need the following basic fact:

[7] We made no effort to optimize this constant factor.

Fact 6.7. *Given two psd matrices ρ, σ, if $|\operatorname{Tr}(\rho) - \operatorname{Tr}(\sigma)| \le \varepsilon/2$ and $\|\rho - \sigma\| \ge \varepsilon$, then*

$$\|\rho/\operatorname{Tr}(\rho) - \sigma/\operatorname{Tr}(\sigma)\|_1 \ge \varepsilon/2\operatorname{Tr}(\rho).$$

Proof. Note that

$$\|\sigma/\operatorname{Tr}(\rho) - \sigma/\operatorname{Tr}(\sigma)\|_1 = \left|\frac{\operatorname{Tr}(\sigma)}{\operatorname{Tr}(\rho)} - 1\right| \le \frac{\varepsilon}{2\operatorname{Tr}(\rho)},$$

so by triangle inequality,

$$\|\rho/\operatorname{Tr}(\rho) - \sigma/\operatorname{Tr}(\sigma)\|_1 \ge \frac{1}{\operatorname{Tr}(\rho)}\|\rho - \sigma\|_1 - \|\sigma/\operatorname{Tr}(\rho) - \sigma/\operatorname{Tr}(\sigma)\|_1 \ge \frac{\varepsilon}{2\operatorname{Tr}(\rho)}.$$

□

6.3 Instance-Near-Optimal Certification

We are ready to prove Theorem 6.1.

Proof of Theorem 6.1. We have that

$$\rho = \sum_{j\in\mathcal{J}} \rho[j,j] + \sum_{j\in\mathcal{J}:j\neq j'} \rho[j,j'] + \rho_{\mathsf{junk}}^{\mathsf{diag}} + \rho_{\mathsf{junk}}^{\mathsf{off}} \qquad \sigma' = \sum_{j\in\mathcal{J}} \sigma[j,j]$$

If $\|\rho - \sigma\|_1 > \varepsilon$, then by triangle inequality,

$$\left\| \sum_{j\in\mathcal{J}} (\rho[j,j] - \sigma[j,j]) + \sum_{j,j'\in\mathcal{J}:j\neq j'} \rho[j,j'] + \rho_{\mathsf{junk}}^{\mathsf{diag}} + \rho_{\mathsf{junk}}^{\mathsf{off}} \right\|_1 = \|\rho - \sigma'\|_1 \ge \varepsilon - \varepsilon^2/20 \ge 9\varepsilon/10$$

and one of four things can happen:

1. $\|\rho_{\mathsf{junk}}^{\mathsf{diag}}\|_1 \ge \varepsilon^2/8$.
2. $\|\rho_{\mathsf{junk}}^{\mathsf{off}}\|_1 \ge \varepsilon/2$,
3. There exists $j \in \mathcal{J}$ for which $\|\rho[j,j] - \sigma[j,j]\|_1 \ge \varepsilon/(10m^2)$
4. There exist distinct $j, j' \in \mathcal{J}$ for which $\|\rho[j,j']\|_1 \ge \varepsilon/(5m^2)$.

Otherwise we would have

$$\|\rho - \sigma'\|_1 \le m \cdot \frac{\varepsilon}{10m^2} + \binom{m}{2} \cdot \frac{\varepsilon}{5m^2} + \frac{\varepsilon^2}{8} + \frac{\varepsilon}{2} = \frac{\varepsilon}{10m} + \frac{\varepsilon(m-1)}{10m} + \frac{3\varepsilon}{4} < 9\varepsilon/10,$$

a contradiction.

It remains to demonstrate how to test whether we are in any of Scenarios 1 to 4.

Lemma 6.8. *$O(\log(1/\delta)/\varepsilon^2)$ copies suffice to test whether $\rho = \sigma$ or whether Scenario 1 holds, with probability $1 - O(\delta)$.*

Proof. We can use the POVM consisting of the projector Π to the principal submatrix indexed by S_{tail}, together with $\mathbb{1} - \Pi$, to distinguish between whether $\operatorname{Tr}(\rho_{\mathsf{junk}}^{\mathsf{diag}}) \ge \varepsilon^2/8$ or whether $\operatorname{Tr}(\rho_{\mathsf{junk}}^{\mathsf{diag}}) \le \varepsilon^2/10$, the latter of which holds if $\rho = \sigma$ by definition of S_{tail}. For this distinguishing task, $O(\log(1/\delta)/\varepsilon^2)$ copies suffice. □

Lemma 6.9. *If Scenario 1 does not hold, then Scenario 2 cannot hold.*

Proof. Suppose Scenario 1 does not hold so that $\|\rho_{\mathsf{junk}}^{\mathsf{diag}}\|_1 < \varepsilon^2/4$. Then by the first part of Lemma 3.15, $\|\rho_{\mathsf{junk}}^{\mathsf{off}}\|_1^2 < (1-\varepsilon^2/4)\cdot\varepsilon^2/4 < \varepsilon^2/4$, a contradiction. □

Lemma 6.10. *$O\left(\|\sigma'\|_{2/5}\operatorname{polylog}(d/\varepsilon)\log(m/\delta)/\varepsilon^2\right)$ copies suffice to test whether $\rho=\sigma$ or whether Scenario 3 holds, with probability $1-O(\delta)$.*

Proof. If $\operatorname{Tr}(\sigma[j,j]) < \varepsilon/(10m^2)$, then to test whether $\rho=\sigma$ or Scenario 3 holds, it suffices to decide whether $\operatorname{Tr}(\rho[j,j]) \geq \operatorname{Tr}(\sigma[j,j]) + \varepsilon/(10m^2)$. We can do this by measuring ρ using the POVM consisting of the projection Π_j to the principal submatrix indexed by S_j, together with $\mathbb{1}-\Pi_j$, for which $O(m^4\log^2(1/\delta)/\varepsilon^2)$ copies suffice to determine this with probability $1-O(\delta)$.

Suppose now that $\operatorname{Tr}(\sigma[j,j]) \geq \varepsilon/(10m^2)$. We can use $O(\log^4(d/\varepsilon)\cdot\log(1/\delta)/\varepsilon^2)$ copies to approximate $\operatorname{Tr}(\rho[j,j])$ to additive error $\varepsilon/(40m^2)$ with probability $1-O(\delta)$ using the same POVM.

If our estimate for $\operatorname{Tr}(\rho[j,j])$ is greater than $\varepsilon/(40m^2)$ away from $\operatorname{Tr}(\sigma[j,j])$, then $\rho\neq\sigma$.

Otherwise, $|\operatorname{Tr}(\rho[j,j]) - \operatorname{Tr}(\sigma[j,j])| \leq \varepsilon/(20m^2)$. Then by Fact 6.7, to determine whether we are in Scenario 3, it suffices to design a tester to distinguish whether the mixed states $\widehat{\rho}[j,j]$ and $\widehat{\sigma}[j,j]$ are equal or ε'-far in trace distance for

$$\varepsilon' \triangleq \frac{\varepsilon}{20m^2\operatorname{Tr}(\sigma[j,j])} = \Theta\left(\frac{\varepsilon}{20m^2 d_j 2^{-j}}\right). \tag{28}$$

Note that if $\widehat{\rho}[j,j]$ and $\widehat{\sigma}[j,j]$ are ε'-far in trace distance, they are at least $\varepsilon'/\sqrt{d_j}$-far in Hilbert-Schmidt. We conclude from Lemma 6.2 that we can distinguish with probability $1-O(\delta)$ between whether $\widehat{\rho}[j,j]$ and $\widehat{\sigma}[j,j]$ are equal or ε'-far in trace distance using $O(d_j^{3/2}\log(1/\delta)/\varepsilon'^2) = O(d_j^{7/2}2^{-2j}\log^4(d/\varepsilon)\log(1/\delta)/\varepsilon^2)$ measurements on the conditional state $\widehat{\rho}[j,j]$. Note that $\operatorname{Tr}(\rho[j,j]) \geq \Omega(\operatorname{Tr}(\sigma[j,j]))$ because $\operatorname{Tr}(\sigma[j,j]) \geq \varepsilon/(10m^2)$ by assumption, so $\operatorname{Tr}(\sigma[j,j]) \geq \Omega(d_j 2^{-j})$. As a result, this tester can make the desired number of measurements on the conditional state by using $O(d_j^{5/2}2^{-j}\log^4(d/\varepsilon)\log(1/\delta)/\varepsilon^2)$ copies of ρ and rejection sampling.

By a union bound over distinct pairs j, j', it therefore takes $O(\log(m/\delta))$ times

$$\sum_{j\in\mathcal{J}} O\left(d_j^{5/2}2^{-j}\log^4(d/\varepsilon)/\varepsilon^2\right) \leq \sum_{j\in\mathcal{J}} O\left(d_j^{5/2}\lambda_j\log^4(d/\varepsilon)/\varepsilon^2\right) \leq O\left(\|\sigma'\|_{2/5}\operatorname{polylog}(d/\varepsilon)/\varepsilon^2\right),$$

copies to test whether Scenario 3 holds, where the last step above follows by Fact 3.18. □

Lemma 6.11. *If Scenario 3 does not hold, then $O\left(\sqrt{d-d'}\|\sigma'\|_{1/2}\log(m/\delta)\operatorname{polylog}(d/\varepsilon)/\varepsilon^2\right)$ copies suffice to test whether $\rho=\sigma$ or whether Scenario 4 holds, with probability $1-O(\delta)$.*

Proof. Fix any $j\neq j'\in\mathcal{J}$ and suppose without loss of generality that $d_j \geq d_{j'}$. Let ρ^* and σ^* denote the matrices obtained by zeroing out all entries of ρ and σ except those in the principal submatrix indexed by $S_j\cup S_{j'}$. Let $\widehat{\rho}^*_{j,j'}$ and $\widehat{\sigma}^*_{j,j'}$ denote these same matrices with trace normalized to 1. For brevity, we will freely omit subscripts.

If $\operatorname{Tr}(\sigma^*) < \varepsilon/(5m^2)$, then $\|\sigma[j,j']\|_1 \leq \varepsilon/(10m^2)$ by the second part of Lemma 3.15. If Scenario 2 holds, then $\|\rho[j,j']\|_1 \geq \varepsilon/(5m^2)$, so by another application of the second part of Lemma 3.15, we would get that $\operatorname{Tr}(\rho^*) \geq 2\varepsilon/(5m^2)$, contradicting the fact that Scenario 1 does not hold.

Suppose now that $\operatorname{Tr}(\sigma^*) \geq \varepsilon/(5m^2)$. As in the proof of Lemma 6.10, we can use $O(\log^4(d/\varepsilon)\cdot\log(1/\delta)/\varepsilon^2)$ copies to approximate $\operatorname{Tr}(\rho^*)$ to within additive error $\varepsilon/(20m^2)$ with probability $1-O(\delta)$.

If our estimate is greater than $\varepsilon/(20m^2)$ away from $\mathrm{Tr}(\sigma[j,j])$ then we know that $\rho \neq \sigma$.

Otherwise, $|\mathrm{Tr}(\rho^*) - \mathrm{Tr}(\sigma^*)| \leq \varepsilon/(10m^2)$, and in particular $\mathrm{Tr}(\rho^*) \geq \Omega(\mathrm{Tr}(\sigma^*))$ as a result. If Scenario 3 holds but Scenario 4 does not, then $\|\rho^* - \sigma^*\| \geq \varepsilon/(5m^2)$. So by Fact 6.7, to determine whether we are in Scenario 2, it suffices to design a tester to distinguish whether the mixed states $\widehat{\rho}^*$ and $\widehat{\sigma}^*$ are equal or ε''-far in trace distance, where

$$\varepsilon'' \triangleq \frac{\varepsilon}{10m^2 \,\mathrm{Tr}(\sigma^*)} = \Theta\left(\frac{\varepsilon}{10m^2} \cdot (d_j 2^{-j} + d_{j'} 2^{-j'})^{-1}\right) \tag{29}$$

Note that if ρ^* and σ^* are ε''-far in trace distance, they are at least $\varepsilon''/\sqrt{d_j}$-far in Hilbert-Schmidt, by the assumption that $d_j \geq d_{j'}$. We conclude from Lemma 6.2 that we can distinguish these two cases using

$$O(\sqrt{d_j} d_{j'} \log(1/\delta)/\varepsilon'^2) = O\left(\sqrt{d_j} d_{j'} (d_j 2^{-j} + d_{j'} 2^{-j'})^2 \log^4(d/\varepsilon) \log(1/\delta)/\varepsilon^2\right)$$

measurements on the conditional state $\widehat{\rho}^*$. Because $\mathrm{Tr}(\rho^*) \geq \Omega(\mathrm{Tr}(\sigma^*)) \geq \Omega(d_j 2^{-j} + d_{j'} 2^{-j'})$, this tester can make the desired number of measurements on the conditional state by using $O\left(\sqrt{d_j} d_{j'} (d_j 2^{-j} + d_{j'} 2^{-j'}) \log^4(d/\varepsilon) \log(1/\delta)/\varepsilon^2\right)$ copies of ρ and rejection sampling.

Summing over $j \neq j' \in \mathcal{J}$ for which $d_j \geq d_{j'}$, we conclude that it takes $O(\log(1/\delta))$ times

$$\begin{aligned}
\sum_{j \neq j' \in \mathcal{J} : d_j \geq d_{j'}} \sqrt{d_j} d_{j'} (d_j 2^{-j} + d_{j'} 2^{-j'}) &\leq \sum_{j,j' \in \mathcal{J} : d_j \geq d_{j'}} d_j^{3/2} d_{j'} 2^{-j} + \sum_{j,j' \in \mathcal{J} : d_j \geq d_{j'}} \sqrt{d_j} d_{j'}^2 2^{-j'} \\
&\leq |\mathcal{J}| \cdot \sum_{j \in \mathcal{J}} d_j^{5/2} 2^{-j} + \left(\sum_{j \in \mathcal{J}} \sqrt{d_j}\right) \left(\sum_{j \in \mathcal{J}} d_j^2 2^{-j}\right) \\
&\leq \mathrm{polylog}(d/\varepsilon) \cdot \left(\|\sigma'\|_{2/5} + \sqrt{d - d'} \cdot \|\sigma'\|_{1/2}\right),
\end{aligned}$$

copies to test whether Scenario 4 holds, where the last step above uses Fact 3.18.

We claim that the above bound is dominated by $O(\log(m/\delta)\,\mathrm{polylog}(d/\varepsilon))\sqrt{d - d'}\|\sigma'\|_{1/2}$. Indeed, note that for any vector $v \in \mathbb{R}^m$,

$$\|v\|_{2/5}^{2/5} = \sum_i v_i^{2/5} \leq \left(\sum_i (v_i^{2/5})^{5/4}\right)^{4/5} \cdot \left(\sum_i 1^5\right)^{1/5} \leq \|v\|_{1/2}^{2/5} \cdot \sqrt{m}^{2/5},$$

as desired. □

Altogether, Lemmas 6.8 to 6.11 allow us to conclude correctness of the algorithm Certify whose pseudocode is provided in Algorithm 2 below. The copy complexity guarantee follows from these lemmas together with Fact 5.27. □

Remark 6.12. *As stated, we are performing measurements in Haar-random bases at various points in* Certify *and in particular the subroutine* BasicCertify*. As Lemma 6.4 and Lemma 6.2 make clear however, we only exploit the first four moments of the Haar measure over the unitary group. As a result, if we were interested in implementing a gate-efficient protocol for state certification, we could have replaced the Haar measure with an approximate 4-design, for which there are a variety of gate-efficient constructions, e.g. [HMMH+ 20].*

Algorithm 2: CERTIFY$(\rho, \sigma, \varepsilon, \delta)$

Input: Copy access to ρ, diagonal density matrix σ, error ε, failure probability δ

Output: YES if $\rho = \sigma$, NO if $\|\rho - \sigma\|_{\mathrm{HS}} > \varepsilon$, with probability $1 - \delta$.

$m \leftarrow \log(10d/\varepsilon^2)$.

Let Π be the projector to the principal submatrix indexed by S_{tail}. `// Scenario 1`

$\mathcal{M} \leftarrow \{\Pi, \mathbb{1} - \Pi\}$.

Measure $O(\log(1/\delta)/\varepsilon^2)$ copies of ρ with the POVM $\mathcal{M}$.

if $\geq (\varepsilon^2/5)$ *fraction of outcomes observed correspond to* Π **then**

  return NO.

for $j \in \mathcal{J}$ **do** `// Scenario 3`

  Let Π_j denote the projection to the principal submatrix indexed by S_j.

  $\mathcal{M}_j \leftarrow \{\Pi_j, \mathbb{1} - \Pi_j\}$.

  Measure $O(\mathrm{polylog}(d/\varepsilon)\log(1/\delta)/\varepsilon^2)$ copies of ρ with the POVM $\mathcal{M}_j$.

  if $\geq (\mathrm{Tr}(\sigma[j,j]) + \varepsilon/(40m^2))$ *fraction of outcomes observed correspond to* Π_j **then**

    return NO.

  else

    Define ε' according to (28).

    $b_j \leftarrow$BASICCERTIFY$(\widehat{\rho}[j,j], \widehat{\sigma}[j,j], \varepsilon', O(\delta/m))$.

    if $b_j =$ NO **then**

      return NO.

for $j, j' \in \mathcal{J}$ *distinct and satisfying* $d_j \geq d_{j'}$ **do** `// Scenario 4`

  Let $\Pi_{j,j'}$ denote the projection to the principal submatrix indexed by $S_j \cup S_{j'}$.

  $\mathcal{M}_{j,j'} \leftarrow \{\Pi_{j,j'}, \mathbb{1} - \Pi_{j,j'}\}$.

  Measure $O(\mathrm{polylog}(d/\varepsilon)\log(1/\delta)/\varepsilon^2)$ copies of ρ with the POVM $\mathcal{M}_{j,j'}$.

  if $\geq (\mathrm{Tr}(\sigma^*_{j,j'}) + \varepsilon/(20m^2))$ *fraction of outcomes observed correspond to* $\Pi_{j,j'}$ **then**

    return NO.

  else

    Define ε'' according to (29).

    $b_{j,j'} \leftarrow$BASICCERTIFY$(\widehat{\rho}^*_{j,j'}, \widehat{\sigma}^*_{j,j'}, \varepsilon'', O(\delta/m^2))$.

    if $b_{j,j'} =$ NO **then**

      return NO.

return YES.

Acknowledgments The authors would like to thank Robin Kothari for helpful discussions at an early stage of this work, as well as Hsin-Yuan Huang for suggesting the approach of lower bounding the likelihood ratio. Part of this work was completed while SC and JL were visiting the Simons Institute for the Theory of Computing.

References

[ACQ21] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *arXiv preprint arXiv:2101.04634*, 2021.

[ADJ+11] Jayadev Acharya, Hirakendu Das, Ashkan Jafarpour, Alon Orlitsky, and Shengjun Pan. Competitive closeness testing. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 47–68. JMLR Workshop and Conference Proceedings, 2011.

[ADJ+12] Jayadev Acharya, Hirakendu Das, Ashkan Jafarpour, Alon Orlitsky, Shengjun Pan, and Ananda Suresh. Competitive classification and closeness testing. In *Conference on Learning Theory*, pages 22–1. JMLR Workshop and Conference Proceedings, 2012.

[AGKE15] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature communications*, 6(1):1–8, 2015.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010.

[ANSV08] Koenraad MR Audenaert, Michael Nussbaum, Arleta Szkoła, and Frank Verstraete. Asymptotic error rates in quantum hypothesis testing. *Communications in Mathematical Physics*, 279(1):251–283, 2008.

[BC09] Stephen M Barnett and Sarah Croke. Quantum state discrimination. *Advances in Optics and Photonics*, 1(2):238–278, 2009.

[BCG19] Eric Blais, Clément L Canonne, and Tom Gur. Distribution testing lower bounds via reductions from communication complexity. *ACM Transactions on Computation Theory (TOCT)*, 11(2):1–37, 2019.

[BCHJ+19] Fernando GSL Brandão, Wissam Chemissany, Nicholas Hunter-Jones, Richard Kueng, and John Preskill. Models of quantum complexity growth. *arXiv preprint arXiv:1912.04297*, 2019.

[BCL20] Sebastien Bubeck, Sitan Chen, and Jerry Li. Entanglement is necessary for optimal quantum property testing. *arXiv preprint arXiv:2004.07869*, 2020.

[BK15] Joonwoo Bae and Leong-Chuan Kwek. Quantum state discrimination and its applications. *Journal of Physics A: Mathematical and Theoretical*, 48(8):083001, 2015.

[BOW19] Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 503–514, 2019.

[Can20] Clément L Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing*, pages 1–100, 2020.

[CCHL21] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. *to appear in FOCS*, 2021.

[Che00] Anthony Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.

[CŚ06] 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.

[DK16] Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694. IEEE, 2016.

[dSLCP11] Marcus P da Silva, Olivier Landon-Cardinal, and David Poulin. Practical characterization of quantum devices without tomography. *Physical Review Letters*, 107(21):210404, 2011.

[FGLE12] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

[FL11] Steven T Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Physical review letters*, 106(23):230501, 2011.

[GLF+10] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[Gol17] Oded Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

[HHJ+17] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.

[HKP21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021.

[HMMH+20] Jonas Haferkamp, Felipe Montealegre-Mora, Markus Heinrich, Jens Eisert, David Gross, and Ingo Roth. Quantum homeopathy works: Efficient unitary designs with a system-size independent number of non-clifford gates. *arXiv preprint arXiv:2002.09524*, 2020.

[IS12] Yuri Ingster and Irina A Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, 2012.

[JHW18] Jiantao Jiao, Yanjun Han, and Tsachy Weissman. Minimax estimation of the l_1 distance. *IEEE Transactions on Information Theory*, 64(10):6672–6706, 2018.

[KRT17] 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.

[MdW16] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. *Theory of Computing*, pages 1–81, 2016.

[MM13] Elizabeth Meckes and Mark Meckes. Spectral measures of powers of random matrices. *Electronic communications in probability*, 18, 2013.

[OW15] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 529–538, 2015.

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the48th Annual ACM symposium on Theory of Computing*, pages 899–912, 2016.

[OW17] Ryan O'Donnell and John Wright. Efficient quantum tomography ii. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, pages 962–974, 2017.

[Pan08] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

[Vor13] Vladislav Voroninski. Quantum tomography from few full-rank observables. *arXiv preprint arXiv:1309.7669*, 2013.

[VV17] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.

[Wu17] Yihong Wu. Lecture notes on information-theoretic methods for high-dimensional statistics. *Lecture Notes for ECE598YW (UIUC)*, 16, 2017.

[Zha06] Fuzhen Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.

A Adaptive Lower Bound

In this section we prove a lower bound against state certification algorithms that use adaptive, unentangled measurements.

Theorem A.1. *There is an absolute constant $c > 0$ for which the following holds for any $0 < \varepsilon < c$.*[8] *Let $\sigma \in \mathbb{C}^{d \times d}$ be a diagonal density matrix. There is a matrix σ^* given by zeroing out the largest entry of σ and at most $O(\varepsilon \log(d/\varepsilon))$ additional mass from σ (see Definition A.2 below), such that the following holds:*

Any algorithm for state certification to error ε with respect to σ using adaptive, unentangled measurements has copy complexity at least

$$\Omega\left(d \cdot d_{\mathsf{eff}}^{1/3} \cdot F(\widehat{\sigma}^*, \rho_{\mathsf{mm}})/(\varepsilon^2 \log(d/\varepsilon))\right).$$

The outline follows that of Section 5. In Section A.1, we describe the procedure by which we remove mass from σ, which will be more aggressive than the one used for our nonadaptive lower bound. As a result, it will suffice to analyze the lower bound instance given in Section 5.3, which we do in Section A.2. For our analysis, we need to check some additional conditions hold for the adaptive lower bound framework of Section 4.3 to apply.

[8] As presented, our analysis yields c within the vicinity of 1/3, but we made no attempt to optimize for this constant.

A.1 Bucketing and Mass Removal

Define $\{S_j\}, \mathcal{J}, S_{\mathsf{sing}}, S_{\mathsf{many}}$ in the same way as in Section 5.1. The way in which we remove mass from σ will be more aggressive than in the nonadaptive setting. We will end up removing up to $O(\varepsilon \log(d/\varepsilon))$ mass (see Fact A.3) as follows:

Definition A.2 (Removing low-probability elements- adaptive lower bound)**.** *Without loss of generality, suppose that $\lambda_1, \ldots, \lambda_d$ are sorted in ascending order according to λ_i. Let $d' \le d$ denote the largest index for which $\sum_{i=1}^{d'} \lambda'_i \le 4\varepsilon$. Let $S_{\mathsf{tail}} \triangleq [d']$.*

Let σ^* denote the matrix given by zeroing out the largest entry of σ and the entries indexed by S_{tail}. It will be convenient to define $\mathcal{J}^*$ to be the buckets for the nonzero entries of σ^*, i.e. the set of $j \in \mathcal{J}$ for which S_j has nonempty intersection with $[d] \backslash S_{\mathsf{tail}}$.

Fact A.3. *There are at most $O(\log(d/\varepsilon))$ indices $j \in \mathcal{J}^*$. As a consequence, $\mathrm{Tr}(\sigma^*) \ge 1 - O(\varepsilon \log(d/\varepsilon))$.*

Proof. For any $i_1 \notin S_{\mathsf{tail}}$ and $i_2 \in S_{\mathsf{tail}}$, we have that $p_{i_1} > p_{i_2}$. In particular, summing over $i_2 \in S_{\mathsf{tail}}$, we conclude that $p_{i_1} \cdot |S_{\mathsf{tail}}| > 4\varepsilon$, so $p_{i_1} > 4\varepsilon/d$. By construction of the buckets S_j, the first part of the claim follows. As in the proof of Fact 5.3, the second part of the claim follows by definition of S_{light}. □

A.2 Analyzing Lower Bound II

We will analyze the sub-problem defined in Section 5.3 and prove the following lower bound:

Lemma A.4. *Fix any $j, j' \in \mathcal{J}^*$ satisfying $d_j \ge d_{j'}$. If $d_j > 1$, then we can optionally take $j = j'$. Suppose $\varepsilon \le d_{j'} \cdot 2^{-j/2-j'/2-1}$. Distinguishing between whether $\rho = \sigma$ or $\rho = \sigma_{\mathbf{W}}$ for $\mathbf{W} \in \mathbb{C}^{d_j \times d_{j'}}$ consisting of Haar-random orthonormal columns (see (15) and (16)), using adaptive unentangled measurements, has copy complexity at least*

$$\Omega\left(\frac{d_j^{1/3} \cdot d_{j'}^2 \cdot 2^{-j'}}{\varepsilon^2}\right).$$

Proof. As in Section 5.3, we will focus on the case where $j \ne j'$, but at the cost of some factors of two, the following arguments easily extend to the construction for $j = j'$ when $d_j > 1$ by replacing S_j and $S_{j'}$ with S_j^1, S_j^2 defined immediately before (16).

We have already verified in Section 5.3 that Conditions 1, 2, and (3) of Assumption 1 are satisfied by $\mathcal{P}$ for $L, \varsigma = O\left(\frac{\varepsilon}{d_{j'} 2^{-j'/2}}\right)$.

It remains to check that $|g_{\mathcal{P}}^{\mathbf{U}}(z)| \le 0.99$ for all z. To this end, recall (18). As the diagonal entries of ρ indexed by S_j (resp. $S_{j'}$) are at least 2^{-j-1} (resp. $2^{-j'-1}$),

$$v_z^\dagger \rho v_z \ge 2^{-j-1}\|v_z^j\|^2 + 2^{-j'-1}\|v_z^{j'}\|^2 \ge 2^{-j/2-j'/2}\|v_z^j\|\|v_z^{j'}\|,$$

so

$$g_{\mathcal{P}}^{\mathbf{U}}(z) \le \frac{\varepsilon}{d_{j'}} \cdot \frac{\|v_z^j\|\|v_z^{j'}\|}{2^{-j/2-j'/2}\|v_z^j\|\|v_z^{j'}\|} \le \frac{\varepsilon}{d_{j'} 2^{-j/2-j'/2}}.$$

In particular, as long as $\varepsilon \le d_{j'} 2^{-j/2-j'/2-1}$, we have the bound $|g_{\mathcal{P}}^{\mathbf{U}}(z)| \le 1/2$.

We can now apply Theorem 4.8 with $\tau = O\left(\frac{\varepsilon^2}{d_j^{1/3} d_{j'}^2 2^{-j'}}\right)$, noting that

$$\exp\left(-\Omega\left(\left\{\frac{d_j\tau^2}{L^2\varsigma^2} \wedge \frac{d\tau}{L^2}\right\}\right)\right) = \exp\left(-\Omega\left(d_j^{1/3}\right)\right),$$

to get that for any adaptive unentangled POVM schedule $\mathcal{S}$, if $p_0^{\leq N}$ is the distribution over outcomes from measuring N copies of σ with $\mathcal{S}$ and $p_1^{\leq N}$ is the distribution from measuring N copies of $\sigma_{\mathbf{U}}$, then

$$\mathrm{KL}\left(p_1^{\leq N} \| p_0^{\leq N}\right) \leq \frac{N\varepsilon^2}{d_j^{1/3} d_{j'}^2 2^{-j'}} + O(N) \cdot \exp\left(-\Omega\left(d_j^{1/3} - \frac{N\varepsilon^2}{d_{j'}^2 2^{-j'}}\right)\right).$$

In particular, if $N = o\left(\frac{d_j^{1/3} d_{j'}^2 2^{-j'}}{\varepsilon^2 \log(d/\varepsilon)}\right)$, then $\mathrm{KL}\left(p_1^{\leq N} \| p_0^{\leq N}\right) = o(1)$ and we get the desired lower bound. □

A.3 Putting Everything Together

Proof of Theorem A.1. As in the proof of Theorem 5.1, we proceed by casework depending on whether $d_j = 1$ for all $j \in \mathcal{J}^*$.

Case 1. *$d_j = 1$ for all $j \in \mathcal{J}^*$.*

The analysis for this case in the nonadaptive setting completely carries over to this setting, because the lower bound from Lemma 5.24 holds even against adaptive POVM schedules. There are two possibilities. If there is a single bucket $j = j(i)$ for which $i \notin S_{\mathsf{tail}}$, then $d_{\mathsf{eff}} = 1$ and $\|\sigma^*\|_{1/2} = O(1)$; for ε smaller than some absolute constant, we have that $\sigma_{i,i} \geq 3/4$ and Lemma 5.24 gives an $\Omega(1/\varepsilon^2)$ lower bound as desired. Otherwise, let j' be the smallest index for which $j' = j(i')$ for some $i' \in \mathcal{J}^*$, and let $j > j'$ be the next smallest index for which $j = j(i)$ for some $i \in \mathcal{J}^*$. Consider the lower bound instance in Section A.2 applied to this choice of j, j'. Provided that $\varepsilon \leq 2^{-j/2-j'/2-1}$, we would obtain a copy complexity lower bound of $\Omega(2^{-j'}/\varepsilon^2) \geq \Omega(\|\sigma^*\|_{1/2}/(\varepsilon^2 \log(d/\varepsilon)))$, where the inequality is by Fact 3.18, and we would be done. On the other hand, if $\varepsilon \geq 2^{-j/2-j'/2-1}$, then because $2^{-j'} > 2^{-j}$, we would conclude that $2^{-j} \leq 2\varepsilon$. In particular, this implies that $\sum_{j'' \in \mathcal{J}^*, i \in S_{j''}: j'' \neq j'} \lambda_i \leq 4\varepsilon$, contradicting the fact that we have removed all buckets of total mass at most 4ε in defining S_{tail}.

Case 2. *$d_j > 1$ for some $j \in \mathcal{J}^*$.*

Let $j_* \triangleq \arg\max_{j \in \mathcal{J}^*} d_j$ and $j'_* \triangleq \arg\max_{j \in \mathcal{J}^*} d_j^2 2^{-j}$. By Lemma 5.19, as long as ε satisfies the bound

$$\varepsilon \leq d_{j'_*} \cdot 2^{-j_*/2-j'_*/2-1}, \tag{30}$$

we have a lower bound of

$$\Omega\left(d_{j_*}^{1/3} \cdot d_{j'_*}^2 \cdot 2^{-j'_*}/\varepsilon^2\right) \geq \Omega\left(d \cdot d_{\mathsf{eff}}^{1/3} \cdot F(\sigma^*, \rho_{\mathsf{mm}})/(\varepsilon^2 \log(d/\varepsilon))\right),$$

where the second step follows by Fact 3.18 and Fact 5.27. Note that because $d_{j_*} > 1$ as we are in Case 2, we do not constrain j_*, j'_* to be distinct necessarily.

But under our assumptions on j, j' and on $\mathcal{J}^*$, (30) must hold:

$$d_{j'} 2^{-j/2-j'/2-1} \geq d_j 2^{-j-1} \geq \varepsilon$$

where the first step follows by the assumption that $j' \triangleq \arg\max_{j \in \mathcal{J}^*} d_j^2 2^{-j}$, and the second by the assumption that every bucket indexed by $\mathcal{J}^*$ has total mass at least 4ε. □

B Deferred Proofs

B.1 Proof of Lemma 4.4

Fix an arbitrary single-copy subproblem $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ for $\mathcal{D}$ the Haar measure over $U(d)$. For any $\mathbf{V} \in U(d)$, define the functions $F_{\mathbf{V}} : U(d) \to \mathbb{R}$ and $G(\mathbf{U})$ by

$$F_{\mathbf{V}}(\mathbf{U}) \triangleq \phi_{\mathcal{M}}^{\mathbf{U},\mathbf{V}} \qquad G(\mathbf{U}) \triangleq \mathop{\mathbb{E}}_{z\sim p_0(\mathcal{M})}[g_{\mathcal{P}}^{\mathbf{U}}(z)^2]^{1/2}.$$

We first show that Condition 1 and 3 from Assumption 1 imply that $F_{\mathbf{V}}$ is mean zero and Lipschitz:

Lemma B.1. *If $\mathcal{P}$ satisfies Assumption 1, then for any $\mathbf{V} \in U(d)$, $F_{\mathbf{V}}$ is $G(\mathbf{V}) \cdot L$-Lipschitz and satisfies $\mathbb{E}_{\mathbf{U}}[F_{\mathbf{V}}(\mathbf{U})] = 0$.*

Proof. For any $\mathbf{U}, \mathbf{U}' \in U(d)$, we have that

$$\begin{aligned} F_{\mathbf{V}}(\mathbf{U}) - F_{\mathbf{V}}(\mathbf{U}') &= \mathop{\mathbb{E}}_{z\sim p_0(\mathcal{M})}[g^{\mathbf{V}}(z) \cdot (g^{\mathbf{U}}(z) - g^{\mathbf{U}'}(z))] \\ &\le \mathop{\mathbb{E}}_{z}[g^{\mathbf{V}}(z)^2]^{1/2} \cdot \mathop{\mathbb{E}}_{z}[(g^{\mathbf{U}}(z) - g^{\mathbf{U}'}(z))^2]^{1/2} \le G(\mathbf{V}) \cdot L \cdot \|\mathbf{U} - \mathbf{U}'\|_{\mathrm{HS}}, \end{aligned}$$

where the first inequality is by Cauchy-Schwarz, and the second is by Condition 3 of Assumption 1.

The second part of the lemma immediately follows from Condition 1 of Assumption 1. □

Next, we use Conditions 2 and 3 of Assumption 1 to bound the expectation and Lipschitzness of G which, combined with Theorem 3.11, implies the following sub-Gaussian tail bound for G:

Lemma B.2. *If $\mathcal{P}$ satisfies Assumption 1, then for any $s > 0$,*

$$\Pr_{\mathbf{U}}[G(\mathbf{U}) > \varsigma + s] \le \exp(-\Omega(ds^2/L^2)).$$

Proof. The function G is L-Lipschitz. To see this, note that for any $\mathbf{U}, \mathbf{V} \in U(d)$,

$$G(\mathbf{U}) - G(\mathbf{V}) \le \mathop{\mathbb{E}}_{z\sim p_0(\mathcal{M})}[(g_{\mathcal{P}}^{\mathbf{U}}(z) - g_{\mathcal{P}}^{\mathbf{V}}(z))^2]^{1/2} \le L \cdot \|\mathbf{U} - \mathbf{V}\|_{\mathrm{HS}},$$

where the first step is triangle inequality and the second is by Condition 3 of Assumption 1.

By Condition 2 and Jensen's, $\mathbb{E}[G(\mathbf{U})] \le \mathbb{E}[g^{\mathbf{U}}(z)^2]^{1/2} \le \varsigma$. The claim then follows by Theorem 3.11. □

We can finally prove Lemma 4.4:

Proof of Lemma 4.4. Note that $\mathbb{E}[\phi^{\mathbf{U},\mathbf{V}}] = 0$ by the second part of Lemma B.1. By the first of Lemma B.1 and Theorem 3.11,

$$\Pr_{\mathbf{U}}[|\phi^{\mathbf{U},\mathbf{V}}| > s] \le \exp\left(-\Omega\left(\frac{ds^2}{L^2 G(\mathbf{V})^2}\right)\right). \tag{31}$$

We can apply Fact 3.20 to the random variable $Y \triangleq G(\mathbf{V})$ by taking the parameters as follows. Set $a \triangleq 2\varsigma$, $\tau(x) \triangleq \exp(-cd(x-\varsigma)^2/L^2)$, and $f(x) \triangleq \exp(-c'ds^2/L^2x^2)$ for appropriate constants $c, c' > 0$. By (31), $\Pr_{\mathbf{U},\mathbf{V}}[|\phi^{\mathbf{U},\mathbf{V}}| > s] \le \mathbb{E}[f(Y)]$, and by Fact 3.20 and Lemma B.2,

$$\mathbb{E}[f(Y)] \le 2\exp\left(-\frac{c'ds^2}{L^2\varsigma^2}\right) + \int_{2\varsigma}^{\infty} \frac{2c'ds^2}{L^2x^3} \cdot \exp\left(-\frac{d}{L^2}\left(c(x-\varsigma)^2 + c's^2/x^2\right)\right) dx$$

Note that for $x \geq 2\varsigma$, by AM-GM,

$$c(x-\varsigma)^2 + c's^2/x^2 \geq \Omega(s(1-\varsigma/x)) \geq \Omega(s),$$

so we can bound

$$\mathbb{E}[f(Y)] \leq 2\exp\left(-\frac{c'ds^2}{L^2\varsigma^2}\right) + \Omega\left(\frac{ds^2}{L^2\varsigma^2}\right)\cdot\exp(-\Omega(ds/L^2)) \leq \exp\left(-\Omega\left(\frac{ds^2}{L^2\varsigma^2}\wedge\frac{ds}{L^2}\right)\right)$$

as claimed. □

B.2 Proof of Theorem 4.8

Here we prove Theorem 4.8 which gives an adaptive lower bound for distinguishing between a state σ and a mixture of alternatives $\{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}}$ under Assumption 1 when $\mathcal{D}$ is the Haar measure over $U(d)$.

In this section, let $\mathcal{D}$ denote the Haar measure over $U(d)$, and suppose that for any POVM $\mathcal{M}$, the single-copy sub-problem $\mathcal{P} = (\mathcal{M}, \sigma, \{\sigma_{\mathbf{U}}\}_{\mathbf{U}\sim\mathcal{D}})$ satisfies Assumption 1.

B.2.1 Additional Notation

We first introduce some notation. Fix an unentangled, adaptive POVM schedule $\mathcal{S}$. Given a transcript of measurement outcomes $z_{<t}$ up to time t, if $\mathcal{M}^{z_{<t}}$ is the POVM used in time step t, then for convenience we will denote $g_{\mathcal{P}}^{\mathbf{U}}$ and $\phi_{\mathcal{P}}^{\mathbf{U},\mathbf{V}}$ by $g_{z_{<t}}^{\mathbf{U}}$ and $\phi_{z_{<t}}^{\mathbf{U},\mathbf{V}}$, $K_{z_{<t}}^{\mathbf{U},\mathbf{V}}$.

Let $p_0^{\leq t}$ (resp. $p_1^{\leq t}$) denote the distribution over transcripts $z_{\leq t}$ of outcomes up to and including time t under measuring σ (resp. $\sigma_{\mathbf{U}}$ for $\mathbf{U}\sim\mathcal{D}$) with the first t steps of $\mathcal{S}$, and define the quantities

$$\Delta(z_{\leq t}) \triangleq \frac{dp_1^{\leq t}}{dp_0^{\leq t}}(z_{\leq t}) \qquad \Psi_{z_{<t}}^{\mathbf{U},\mathbf{V}} \triangleq \prod_{i=1}^{t-1}(1+g_{z_{<i}}^{\mathbf{U}})(1+g_{z_{<i}}^{\mathbf{V}}),$$

where $\Delta(\cdot)$ is given by the Radon-Nikodym derivative.

B.2.2 Helper Lemmas

We will need the following helper lemmas. The first gives a *lower bound* on the likelihood ratio between $p_1^{\leq t}$ and $p_0^{\leq t}$.

Lemma B.3 (Implicit in Lemma 6.2 of [BCL20])**.** *Under the hypotheses of Theorem 4.8, for any transcript $z_{\leq t}$, $\Delta(z_{\leq t}) \geq \exp(-4\varsigma^2 t)$.*

Proof. By convexity of the exponential function and the fact that $1 + g_{z_{<t}}^{\mathbf{U}}(z_t) > 0$ for all $\mathbf{U}, t, z_t$,

$$\Delta(z_{<t}) \geq \prod_{i=1}^{t-1}\exp\left(\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{D}}[\ln(1+g_{z_{<i}}^{\mathbf{U}}(z_i))]\right).$$

For any $i < t$ we have that

$$\begin{aligned}\exp\left(\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{D}}[\ln(1+g_{z_{<i}}^{\mathbf{U}}(z_i))]\right) &\geq \exp\left(\mathop{\mathbb{E}}_{\mathbf{U}}[g_{z_{<i}}^{\mathbf{U}}(z_i) - 4g_{z_{<i}}^{\mathbf{U}}(z_i)^2]\right)\\ &\geq \exp\left(-4\varsigma^2\right),\end{aligned}$$

where the first step follows by the elementary inequality $\ln(x) \geq x - 4x^2$ for all $x \in [-0.99, 0.99]$ and the fact that $|g_{z_{<t}}^{\mathbf{U}}(z_t)| \leq 0.99$ by hypothesis, and the second step follows by Conditions 1 and 2 of Assumption 1. □

The next lemma gives a bound on the expectation of $(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}})^2$.

Lemma B.4. *Under the hypotheses of Theorem 4.8,* $\mathbb{E}_{z_{<t},\mathbf{U},\mathbf{V}}\left[(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}})^2\right] \le \exp(O(t\varsigma^2))$.

To prove this, it will be convenient to define the following for any ℓ-copy sub-problem corresponding to POVM $\mathcal{M}$

$$K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}} \triangleq \mathop{\mathbb{E}}_{z\sim p_0(\mathcal{M})}\left[\left(g^{\mathbf{U}}_{\mathcal{P}}(z) + g^{\mathbf{V}}_{\mathcal{P}}(z)\right)^2\right]$$

and first show the following:

Lemma B.5. *Under the hypothesis of Theorem 4.8,* $\mathbb{E}_{\mathbf{U},\mathbf{V}}\left[\left(1+\gamma K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}}\right)^t\right] \le \exp(O(\gamma t\varsigma^2))$ *for any absolute constant* $\gamma > 0$ *and any* $t = o(d/L^2)$.

Proof. By the elementary inequality $(a+b)^2 \le 2a^2+2b^2$, we have that $K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}} \le G(\mathbf{U})^2 + G(\mathbf{V})^2$. By Lemma B.2, we immediately get that $\Pr_{\mathbf{U}}\left[K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}} > (\mathbb{E}[G(\mathbf{U})]+s)^2\right] \le \exp(-ds^2/L^2)$. Applying the inequality again allows us to lower bound the left-hand side by $\Pr_{\mathbf{U}}\left[K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}} > 2\,\mathbb{E}[G(\mathbf{U})]^2 + 2s^2\right]$, so we conclude that

$$\Pr_{\mathbf{U}}\left[K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}} > 2\,\mathbb{E}[G(\mathbf{U})]^2 + s\right] \le \exp(-ds/2L^2).$$

We can apply Fact 3.20 to the random variable $Z \triangleq K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}}$ and the function $f(Z) \triangleq (1+\gamma Z)^t$ to conclude that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\left(1+\gamma\cdot K^{\mathbf{U},\mathbf{V}}_{\mathcal{P}}\right)^t\right] &\le 2(1+2\gamma\,\mathbb{E}[G(\mathbf{U})]^2)^t + \int_0^\infty \gamma t(1+\gamma x)^{t-1}\cdot e^{-x\cdot d/2L^2}\,dx \\
&\le 2(1+2\gamma\,\mathbb{E}[G(\mathbf{U})]^2)^t + \gamma t\int_0^\infty e^{-x(d/2L^2-\gamma(t-1))}\,dx \\
&\le 2(1+2\gamma\,\mathbb{E}[G(\mathbf{U})]^2)^t + \frac{\gamma t}{d/2L^2-\gamma(t-1)} \le \exp(O(t\gamma\,\mathbb{E}[G(\mathbf{U})]^2)),
\end{aligned}$$

where in the last two steps we used that $t = o(d/L^2)$ to ensure that the integral is bounded and that the second term in the final expression is negligible. □

We can now prove Lemma B.4:

Proof of Lemma B.4. As $g^{\mathbf{V}}_{z_{<t-1}}(z) \le O(1)$, we know that for any constant $a, b \ge 2$,

$$\mathop{\mathbb{E}}_{z\sim\Omega(\mathcal{M}^{z_{<t-1}})}\left[g^{\mathbf{U}}_{z_{<t-1}}(z)^a\cdot g^{\mathbf{V}}_{z_{<t-1}}(z_t)^b\right] \le \frac{1}{4}\,\mathbb{E}_z[g^{\mathbf{U}}_{z_{<t-1}}(z)^2],$$

so we conclude that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{z\sim p_0(\mathcal{M}^{z_{<t-1}})}\left[(1+g^{\mathbf{U}}_{z_{<t-1}}(z))^c(1+g^{\mathbf{V}}_{z_{<t-1}}(z))^c\right] \\
&\qquad \le 1 + O_c\left(\mathbb{E}_z[g^{\mathbf{U}}_{z_{<t-1}}(z)^2]\right) + O_c\left(\mathbb{E}_z[g^{\mathbf{V}}_{z_{<t-1}}(z)^2]\right) + O_c\left(\phi^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right) \le 1 + C(c)\cdot K^{\mathbf{U},\mathbf{V}}_{z_{<t-1}} \qquad (32)
\end{aligned}$$

for some absolute constant $C(c) > 0$, where the last step follows by AM-GM. For $\alpha_i \triangleq 2 \cdot \left(\frac{t-1}{t-2}\right)^i$, we have that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{z_{<t},\mathbf{U},\mathbf{V}}\left[\left(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right)^{\alpha_i}\right] \\
&\quad\le \mathop{\mathbb{E}}_{z_{<t-1},\mathbf{U},\mathbf{V}}\left[\left(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right)^{\alpha_i}\cdot\left(1+C(\alpha_i)\cdot K^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right)\right] &(33)\\
&\quad\le \mathop{\mathbb{E}}_{z_{<t-1},\mathbf{U},\mathbf{V}}\left[\left(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right)^{\alpha_i(t-1)/(t-2)}\right]^{(t-2)/(t-1)}\cdot \mathop{\mathbb{E}}_{z_{<t-1},\mathbf{U},\mathbf{V}}\left[\left(1+C(\alpha_i)\cdot K^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right)^{t-1}\right]^{1/(t-1)} &(34)\\
&\quad\le \mathop{\mathbb{E}}_{z_{<t-1},\mathbf{U},\mathbf{V}}\left[\left(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right)^{\alpha_{i+1}(t-1)/(t-2)}\right]\cdot \mathop{\mathbb{E}}_{z_{<t-1},\mathbf{U},\mathbf{V}}\left[\left(1+C(\alpha_i)\cdot K^{\mathbf{U},\mathbf{V}}_{z_{<t-1}}\right)^{t-1}\right]^{1/(t-1)}.
\end{aligned}$$

where (33) follows by (32), and (34) follows by Holder's. Unrolling this recurrence, we conclude that

$$\begin{aligned}
\mathop{\mathbb{E}}_{z_{<t},\mathbf{U},\mathbf{V}}\left[\left(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right)^2\right] &\le \prod_{i=1}^{t-1}\mathop{\mathbb{E}}_{z_{<i},\mathbf{U},\mathbf{V}}\left[\left(1+C(\alpha_{t-1-i})\cdot K^{\mathbf{U},\mathbf{V}}_{z_{<i}}\right)^{t-1}\right]^{1/(t-1)}\\
&\le \prod_{i=1}^{t-1}\mathop{\mathbb{E}}_{z_{<i},\mathbf{U},\mathbf{V}}\left[\left(1+C(2e)\cdot K^{\mathbf{U},\mathbf{V}}_{z_{<i}}\right)^{t-1}\right]^{1/(t-1)}, &(35)\\
&\le \sup_{\mathcal{M}}\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\left(1+O(K^{\mathbf{U},\mathbf{V}}_{\mathcal{M}})\right)^{t-1}\right]
\end{aligned}$$

where (35) follows by the fact that for $1 \le i \le t-1$, $\alpha_{t-1-i} \le 2\left(1+\frac{1}{t-2}\right)^{t-2} \le 2e$, and the supremum in the last step is over all POVMs $\mathcal{M}$. The lemma then follows from Lemma B.5. □

B.2.3 Putting Everything Together

The key inequality used in [BCL20] is the following consequence of the chain rule for KL:

Lemma B.6 (Lemma 6.1, [BCL20])**.**

$$KL\left(p_1^{\le N}\|p_0^{\le N}\right) \le \sum_{t=1}^{N} Z_t \quad \textit{for} \quad Z_t \triangleq \mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}\left[\frac{1}{\Delta(z_{<t})}\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\phi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right]\right].$$

We now have all the ingredients to complete the proof of Theorem 4.8.

Proof of Theorem 4.8. Given transcript $z_{<t}$ and $\mathbf{U},\mathbf{V}\sim\mathcal{D}$, let $\mathbb{1}\left[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)\right]$ denote the indicator of whether $\left|\phi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right| > \tau$; note that by Lemma 4.4, this event happens with probability at most $\xi(\tau)$, where

$$\xi(s) \triangleq \exp\left(-\Omega\left(\frac{ds^2}{L^2\varsigma^2}\wedge\frac{ds}{L^2}\right)\right).$$

We have that

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\phi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right] &= \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\phi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\left(\mathbb{1}[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)]+\mathbb{1}[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)^c]\right)\right]\\ &\le \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\mathbb{1}[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)]\right]+\tau\cdot\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\mathbb{1}[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)^c]\right]\\ &\le \cdot\underbrace{\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\cdot\mathbb{1}[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)]\right]}_{\text{(B)}_{z_{<t}}}+\tau\cdot\underbrace{\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{V}}\left[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right]}_{\text{(G)}_{z_{<t}}},\end{aligned}$$

where in the second step we used the assumption that $|g^{\mathbf{U}}_{z_{<t}}(z_t)|\le 0.99$ for all z_t to conclude that $\phi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\le 1$. Note that for any transcript $z_{<t}$, $\Delta(z_{<t})^2=\mathbb{E}_{\mathbf{U},\mathbf{V}}[\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}]=\text{(G)}_{z_{<t}}$, so by this and the fact that the likelihood ratio between two distributions always integrates to 1,

$$\mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}\left[\frac{1}{\Delta^{(t-1)}(z_{<t})}\cdot\text{(G)}_{z_{<t}}\right]=\mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}[\Delta^{(t-1)}(z_{<t})]=1. \tag{36}$$

Recalling the definition of Z_t in Lemma B.6, we conclude that

$$\begin{aligned}Z_t &\le \mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}\left[\frac{1}{\Delta^{(t-1)}(z_{<t})}\cdot\text{(B)}_{z_{<t}}\right]+\tau\cdot\mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}\left[\frac{1}{\Delta^{(t-1)}(z_{<t})}\cdot\text{(G)}_{z_{<t}}\right]\\ &\le \exp(4t\varsigma^2)\mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}\left[\text{(B)}_{z_{<t}}\right]+\tau,\end{aligned}$$

where the second step follows by Lemma B.3 and (36).

To upper bound $\mathbb{E}_{z_{<t}\sim p_0^{\le t-1}}[\text{(B)}_{z_{<t}}]$, apply Cauchy-Schwarz to get

$$\begin{aligned}\mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1}}\left[\text{(B)}_{z_{<t}}\right] &\le \mathop{\mathbb{E}}_{z_{<t}\sim p_0^{\le t-1},\mathbf{U},\mathbf{V}}\left[\left(\Psi^{\mathbf{U},\mathbf{V}}_{z_{<t}}\right)^2\right]^{1/2}\cdot\Pr_{z_{<t}\sim p_0^{\le t-1},\mathbf{U},\mathbf{V}}\left[\mathcal{E}^{\mathbf{U},\mathbf{V}}_{z_{<t}}(\tau)\right]^{1/2}\\ &\le \exp(O(t\varsigma^2))\cdot\xi(\tau),\end{aligned}$$

where the second step follows by Lemma 4.4 and Lemma B.4. Invoking Lemma B.6 concludes the proof. □

B.3 Proof of Fact 5.16

Proof. We may assume $s<m+n$ (otherwise obviously $b=n$). Assume to the contrary that $\sum_{i=1}^{b+1}v_id_i\le\varepsilon$. We proceed by casework based on whether $w_{s'+1}=u_{a+1}$ or $w_{s'+1}=v_{b+1}$.

If $w_{s'+1}=u_{a+1}$, then

$$3\varepsilon<\sum_{i=1}^{s+1}w_id^*_i=\sum_{i=1}^{a+1}u_i+\sum_{i=1}^{b}v_id_i\le\sum_{i=1}^{a+1}v_{b+1}\cdot 2^{1-i}+\sum_{i=1}^{b}v_i\le 2\varepsilon+\sum_{i=1}^{b}v_id_i,$$

where in the first step we used maximality of s, in the third step we used that $u_{a+1}\le v_{b+1}$ and that $u_{i+1}\ge 2u_i$ for all i, and in the last step we used that $v_{b+1}\le\sum_{i=1}^{b+1}v_id_i\le\varepsilon$. From this we conclude that $\sum_{i=1}^{b}v_id_i>\varepsilon$, a contradiction.

If $w_{s'+1}=v_{b+1}$, the argument is nearly identical. We have

$$3\varepsilon<\sum_{i=1}^{s+1}w_id^*_i=\sum_{i=1}^{a}u_i+\sum_{i=1}^{b+1}v_id_i\le\sum_{i=1}^{a}v_{b+1}\cdot 2^{1-i}+\sum_{i=1}^{b+1}v_id_i\le 2\varepsilon+\sum_{i=1}^{b+1}v_i,$$

where in the first step we again used maximality of s, in the third step we used that $u_a \leq v_{b+1}$ and $u_{i+1} \geq 2u_i$ for all i, and in the last step we used that $v_{b+1} \leq \sum_{i=1}^{b+1} v_i d_i \leq \varepsilon$. From this we conclude that $\sum_{i=1}^{b} v_i d_i > \varepsilon$, a contradiction. □