

Sharp bounds for perfect quantum state classification beyond antidistinguishability

Nathaniel Johnston^{*} Benjamin Lovitz[†] Vincent Russo[‡] Jamie Sikora[§]

September 16, 2026

Abstract

A multiset of pure quantum states is said to be *k-learnable* if there is a measurement strategy that always narrows an unknown sample drawn from the list down to one of at most k candidates. The parameter k interpolates between distinguishability and antidistinguishability, and provides a unified framework for partial state identification. We prove two universal, and optimal, Gram-matrix criteria for k -learnability: a Frobenius-norm sufficient condition and an entrywise- ℓ_1 necessary condition. We apply them to derive explicit learnability and copy-complexity guarantees for several well-known sets of states including SIC-POVMs, mutually unbiased bases, and stabilizer states. We further apply our results to zero-error mutation detection problems such as anomaly detection and changepoint detection.

1 Introduction

A central problem in quantum information is to extract classical information from quantum states. The most basic version asks: given a single copy of an unknown pure state drawn from a known finite multiset $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$, can a measurement be designed that reliably reveals which state was drawn? When such a measurement always succeeds in this task, we say that S is *distinguishable*; otherwise some pairs of states inevitably get confused for one another. Between the two extremes of perfect distinguishability and complete indistinguishability lies a rich landscape of partial-information tasks: for instance, one may ask not for the identity of the state but only to *rule out* some candidates, leading to the notion of *antidistinguishability* [CFS02, PBR12, BJOP14, Lei14, HK18, HB20, RS23, MNW24, JRS25, JLRS25]. The concept was originally introduced by Caves, Fuchs, and Schack under the name *post-Peierls incompatibility* [CFS02] and an equivalent formulation in terms of *perfect quantum state exclusion* was given by Bandyopadhyay, Jain, Oppenheim, and Perry [BJOP14]. It has also played a central role in the celebrated PBR theorem [PBR12].

In this paper, we study a common refinement of these notions called *k-learnability*, which we introduced in [JLRS25]. Informally, S is *k-learnable* if a measurement can be designed whose outcome always narrows the identity of the unknown state down to one of at most k candidates. The case $k = 1$ is equivalent to the task of distinguishability and the case $k = n - 1$ is equivalent to the task of antidistinguishability. This task for intermediate values of k is a discrete interpolation

^{*}Department of Mathematics & Computer Science, Mount Allison University. njohnston@mta.ca

[†]Department of Computer Science and Software Engineering, Concordia University. benjamin.lovitz@concordia.ca

[‡]Unitary Foundation and IonQ. vincentrusso1@gmail.com

[§]Department of Computer Science, Virginia Tech. sikora@vt.edu

between the two. Perhaps surprisingly, the states need not be pairwise orthogonal to accomplish this task perfectly: there exist non-orthogonal pure states that are k -learnable for $k \geq 2$. Understanding the conditions under which a multiset of pure states is k -learnable is the focus of this work.

Definition 1. Let $n \geq 2$ be a positive integer and let $k \in \{1, \dots, n\}$. A multiset of pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ is k -learnable if there exists a POVM

$$\{M_I : I \subseteq \{1, 2, \dots, n\}, |I| = k\}, \quad (1)$$

such that

$$\langle \psi_i | M_I | \psi_i \rangle = 0 \quad \text{whenever } i \notin I. \quad (2)$$

In other words, if $|\psi_i\rangle$ is measured with such a POVM, only the subsets of size k containing i could occur.

1.1 Main results

We now discuss our main results and their extensions, corollaries, and applications.

Optimal bounds. We study this new classification task by asking two complementary questions.

What conditions on the states guarantee k -learnability? What conditions on the states forbid k -learnability?

To this end, we provide two complementary bounds (proofs in the appendix).

Theorem 2. Let $n \geq 2$ be an integer, $k \in \{1, 2, \dots, n\}$, and let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$. If

$$\sum_{i,j=1}^n |\langle \psi_i | \psi_j \rangle|^2 \leq \frac{n^2}{n-k+1}, \quad (3)$$

then S is k -learnable. In particular, if

$$|\langle \psi_i | \psi_j \rangle| \leq \sqrt{\frac{k-1}{(n-1)(n-k+1)}} \quad \text{for all } i \neq j \quad (4)$$

then S is k -learnable.

This theorem shows that if the inner products are not collectively *large*, then we can guarantee k -learnability. Of significance, we do not require pairwise orthogonality when $k \geq 2$. The left-hand side of (3) is called the *frame potential* of S [BF03]. The case $S \subset \mathbb{C}^{n-k+1}$ is particularly interesting. In this case, it is known that the right-hand side of (3) is the smallest possible frame potential of any set of n states [BF03] (see also the *Welch bound* [Wel74]). In particular, our theorem shows that any set of n pure states in $\mathbb{C}^{n-k+1}$ attaining the minimum possible frame potential is k -learnable.

On the other hand, if the states are *too close together*, they cannot be k -learnable.

Theorem 3. Let $n \geq 2$ be an integer, $k \in \{1, 2, \dots, n\}$, and let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$. If

$$\sum_{i,j=1}^n |\langle \psi_i | \psi_j \rangle| > nk, \quad (5)$$

then S is not k -learnable. In particular, if

$$|\langle \psi_i | \psi_j \rangle| > \frac{k-1}{n-1} \quad \text{for all } i \neq j \quad (6)$$

then S is not k -learnable.

Together, the above two bounds give explicit criteria for when pure states must be or cannot be k -learnable.

A remark about multisets. While we sometimes describe the task of state discrimination as “determining which state was sent”, a more accurate description is to “determine the index of the state that was sent.” This is because we allow multiple copies of a single state, or those differing by only a global phase, to be within the multiset. For example, for the multiset $\{|\psi_1\rangle = |0\rangle, |\psi_2\rangle = |0\rangle\}$, determining which state was sent is easy, but determining the index of the state that was sent is hard.

A remark about the dimensions of the pure states. We note that sometimes we specify the dimension of the pure states that we wish to k -learn and sometimes it is omitted, like in Theorem 3. Many of our results rely only on the Gram matrix G of the multiset of states which is defined as

$$G_{i,j} = \langle \psi_i | \psi_j \rangle. \quad (7)$$

When a multiset of states is described only by its Gram matrix, this does not fully specify the states (although they are guaranteed to exist). We note that the rank of the Gram matrix is equal to the minimum dimension of any collection of vectors that have that Gram matrix, so there is an implicit bound on the dimension required to realize such states.

Simplex states and 2-learnability. Since perfect distinguishability is a strict condition (pairwise orthogonality), the next best scenario is being 2-learnable. In other words, one can correctly identify the state with 2 guesses, without error. It turns out that the case of $k = 2$ is interesting in light of the two previously mentioned results. From Theorem 2, we have that $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$ is 2-learnable when

$$|\langle \psi_i | \psi_j \rangle| \leq \frac{1}{n-1} \quad (8)$$

for all $i \neq j$ and from Theorem 3, we have that S cannot be 2-learnable when

$$|\langle \psi_i | \psi_j \rangle| > \frac{1}{n-1} \quad (9)$$

for all $i \neq j$. Thus, when $|\langle \psi_i | \psi_j \rangle| = \frac{1}{n-1}$ for $i \neq j$ we have a multiset of states that is 2-learnable, but just barely.

Consider the multiset of states $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subseteq \mathbb{C}^{n-1}$ which satisfies

$$\langle \psi_i | \psi_j \rangle = -\frac{1}{n-1} \quad \text{for all } i \neq j \quad (10)$$

called the *simplex states*. Note that we explicitly mentioned the dimension as $n - 1$ since the Gram matrix has rank $n - 1$. Thus, we have n states in $n - 1$ dimensions that are not even linearly independent but still *spread out enough* to be 2-learnable.

The simplex states saturate all the inequalities in both of Theorems 2 and 3 with equality. If *any* inner product were to increase in magnitude (by possibly increasing the dimension), then they would cease to be 2-learnable, but if *any* inner product were to decrease in magnitude then 2-learnability is still guaranteed.

Simplex states generalize the *trine states* when $n = 3$ and the *tetrahedral states* when $n = 4$, both depicted below.

We now state a result about 2-learnability which is proved in Appendix A.1.

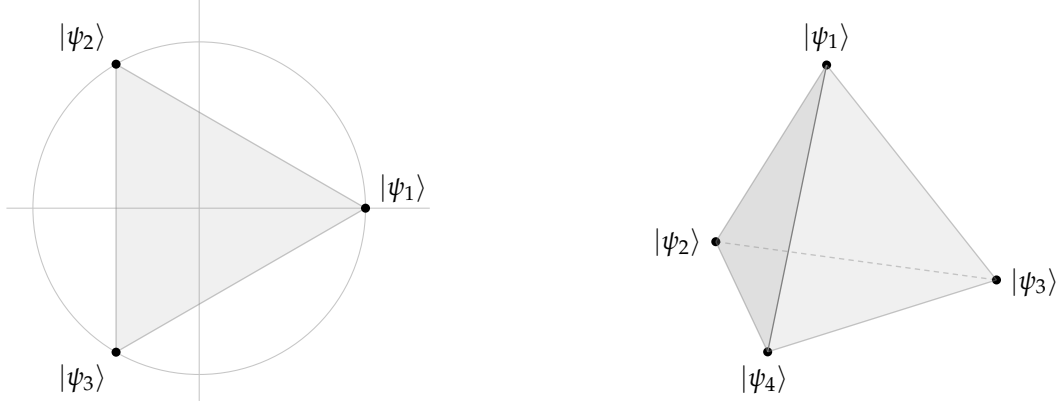


Figure 1: Simplex states for $n = 3$ (left) and $n = 4$ (right), drawn as unit vectors in $\mathbb{R}^{n-1}$. Explicitly, $|\psi_1\rangle = |0\rangle$ and $|\psi_{2,3}\rangle = -\frac{1}{2}|0\rangle \pm \frac{\sqrt{3}}{2}|1\rangle$ when $n = 3$, and $|\psi_i\rangle = \frac{1}{\sqrt{3}}(\pm|0\rangle \pm |1\rangle \pm |2\rangle)$ with an even number of minus signs when $n = 4$. Both multisets are 2-learnable and saturate Theorems 2 and 3 at $k = 2$.

Proposition 4. Let $n \geq 2$ be an integer and let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$. For each $i \in \{1, 2, \dots, n\}$, define

$$R_i := \sum_{j=1}^n |\langle \psi_i | \psi_j \rangle|. \quad (11)$$

The following implications hold:

$$R_i \leq 2 \text{ for all } i \implies S \text{ is 2-learnable} \implies \sum_{i=1}^n R_i \leq 2n. \quad (12)$$

In particular, if $R := R_i$ does not depend on i , then S is 2-learnable if and only if $R \leq 2$.

In many examples of pure states that we consider, we have $R_1 = \dots = R_n$ and thus the above proposition pins down exactly their 2-learnability.

Sharpness of the bounds in Theorems 2 and 3 (Sections 2 and 3). While Theorems 2 and 3 are true, one may ask if these bounds are too aggressive, that is, if the right-hand side of the inequalities can be improved. We now discuss the optimality of these bounds.

Inequality (3) in Theorem 2 is saturated by an explicit family of rank- $(n - k + 1)$ Fourier-type Gram matrices (see Section 3), and we exhibit an explicit ε -deformation of this family that ceases to be k -learnable for all $\varepsilon \in (0, 1)$. The optimality of Inequality (4) is left as an open problem.

Both bounds in Theorem 3 are optimal which can be seen by studying pure states with equal pairwise overlap. This parameterized multiset of pure states is discussed in detail in Section 2 as well as why they are optimal for Theorem 3.

Packing bounds (Section 4). Given that this new perfect classification task is closely tied to how close together or far apart the vectors are, this raises the question:

Given $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ which is k -learnable, what are the relationships between the parameters n , k , and d ?

Such a bound can take the form of a *packing inequality*; mathematically, it describes how many unit vectors forming a k -learnable multiset can be packed into a d -dimensional space. We show that

$$n \leq dk \quad (13)$$

must hold in general. Moreover, this bound can be improved to

$$n \leq \frac{d(k+1)}{2} \quad (14)$$

when no two states are global phases of one another. We prove that both of these bounds are tight via matching constructions.

Copy complexity (Section 5). Intuitively, if instead of being given $|\psi_i\rangle$ to classify one is given c copies of it, i.e., $|\psi_i\rangle^{\otimes c}$, then this makes the task of classification easier. Mathematically, taking additional copies decreases the magnitude of every overlap that lies strictly between zero and one. In particular, we ask:

Given $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$ such that $\{|\psi_1\rangle^{\otimes c}, |\psi_2\rangle^{\otimes c}, \dots, |\psi_n\rangle^{\otimes c}\}$ is k -learnable, what are the relationships between the parameters n, k, c , and S ?

One way to interpret this question comes from the *copy complexity* perspective; what is the minimum c (if it exists) such that a multiset S becomes k -learnable? We use Theorem 2 and Proposition 4 to bound the copy complexity of several well-known sets of states, notably:

Theorem 5 (Informal). *For prime qudit dimension d , the copy complexity of 2-learning the set of m -qudit stabilizer states is $\Theta(m)$.*

See Corollary 27 for a more precise statement. By using m additional copies, one can learn an unknown m -qudit stabilizer state with failure probability at most d^{-m} (see Remark 28 for more details), resulting in an optimal $O(m)$ copy complexity for learning stabilizer states. This reproduces the information-theoretic $O(m)$ copy complexity obtained in [AG08, Mon17, ADIS25, CCLW26], with the caveat that our result only applies to prime-dimensional qudits. In contrast to these works, our 2-learning measurement may not be efficiently implementable. Despite this drawback, an advantage to our result is that the 2-learning approach additionally yields a distinguished pair of stabilizer states guaranteed to contain the unknown state.

More generally, we upper bound the copy complexity for k -learning stabilizer states (Corollary 29). We also give upper bounds on the copy complexity of k -learning arbitrary sets of states (Corollary 18), SIC-POVMs (Corollary 21), and mutually unbiased bases (MUBs, Corollary 23). In the case of 2-learning, we determine the copy complexity for SIC-POVMs and MUBs exactly (Corollaries 20 and 22). In both cases, the copy complexity of 2-learning is at most 4.

We note that our bounds are in stark contrast to the task of perfect distinguishability; a multiset of states that is not perfectly distinguishable will never become perfectly distinguishable by increasing the number of copies. For $k \geq 2$, however, a multiset that is not k -learnable may become k -learnable when sufficiently many copies are provided; Corollary 18 shows this happens whenever the states are pairwise distinct up to global phase.

Application: Zero-error quantum state mutation detection problems (Section 6). There are many applications for which identifying quantum states is of central importance. As an application of our bounds, we revisit the task of anomaly detection [LR18, LSMT24] and changepoint

detection [SBC⁺16, SCMT17, MSU23]. In anomaly detection, we have states of the form

$$|\psi\rangle^{\otimes i-1} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes n-i} \quad (15)$$

and the goal is to identify the location of the state $|\phi\rangle$ (called the anomaly). In changepoint detection, we have states of the form

$$|\phi\rangle|\phi\rangle|\phi\rangle \cdots |\phi\rangle|\psi\rangle \cdots |\psi\rangle \quad (16)$$

and the goal is to identify the first location when the states changed (called the changepoint). In both settings, the goal is to locate a deviation from an otherwise regular sequence of states.

In this paper, we study these problems via the following problem:

When can we narrow the location of the anomaly or changepoint down to k choices, and without error, as a function of the overlap $\langle\psi|\phi\rangle$?

As an example result, we show that the multiset $\{|\psi\rangle^{\otimes i-1} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes n-i} : i \in \{1, 2, \dots, n\}\}$ is k -learnable if and only if

$$|\langle\phi|\psi\rangle| \leq \sqrt{\frac{k-1}{n-1}}. \quad (17)$$

See Theorem 30 for a precise statement. Similar results, including a study of the *multi-anomaly detection* problem [LSMT24], appear in Section 6.

Typical behaviour (Section 7). We now consider how learnable randomly-generated pure states are. Since randomly generated states in a large dimension are almost orthogonal, we expect a trade-off: as the dimension d increases or the number of states n decreases, learnability should increase, and as d decreases or n increases learnability should decrease. We start with a negative result in the regime where d is small or n is large:

Proposition 6. *For fixed $\delta \in (0, 1)$, independent Haar-random pure states $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle \in \mathbb{C}^d$ are not k -learnable with probability at least $1 - \delta$ whenever*

$$d \leq C_\delta \min\{n, (n/k)^2\} \quad (18)$$

for $k \in \{1, \dots, n-1\}$, where $C_\delta > 0$ is a constant that depends only on δ .

See Proposition 33 for a precise statement. Conversely, if n is small or d is large, the randomly generated pure states will be k -learnable with high probability:

Proposition 7. *For fixed $\delta \in (0, 1)$, independent Haar-random pure states $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle \in \mathbb{C}^d$ are k -learnable with probability at least $1 - \delta$ whenever*

$$d \geq C_\delta(n(n-k+1)/k) \quad (19)$$

for $k \in \{2, \dots, n\}$, where $C_\delta > 0$ is a constant that depends only on δ . In particular, $d \leq C_\delta(n^2/k)$ suffices when $k \leq n/2$.

See Proposition 34 for a precise statement.

Methods. Operationally, k -learnability is a property of the existence of measurement strategies. Algebraically, however, it is a property of the Gram matrix. To specify this property, we define what it means for a matrix to be k -incoherent.

Definition 8 (k -incoherence [RBC⁺18]). Let M be an $n \times n$ positive semidefinite matrix. We say that M is k -incoherent if it can be written as a sum of positive semidefinite matrices, each of which is supported only on a $k \times k$ principal submatrix. In other words, M has a decomposition

$$M = \sum_{\ell} M_{\ell}, \quad (20)$$

where each M_{ℓ} is positive semidefinite and has $n - k$ rows that are identically 0 (and similarly for those columns).

Remark 9. We note that the set of k -incoherent matrices is sometimes called the set of matrices with factor width at most k [BCPT05]. Also, the term “ k -incoherent” is sometimes reserved for density matrices (i.e., positive semidefinite matrices with trace 1); we do not make this restriction.

In the prior work [JLRS25, Lemma 9], we showed that determining the learnability of a multiset of pure states is exactly tied to the incoherence of its Gram matrix:

Proposition 10. A multiset of pure states $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ is k -learnable if and only if its Gram matrix is k -incoherent.

Many of our results use this algebraic characterization to determine k -learnability.

2 Sharpness of Theorem 3 via equiangular states

We now discuss a particular parameterized multiset of pure states which all have equal inner product with each other, called *equiangular states*, and are defined implicitly via the $n \times n$ Gram matrix:

$$G_n(\gamma) = \begin{bmatrix} 1 & \gamma & \gamma & \cdots & \gamma \\ \gamma & 1 & \gamma & \cdots & \gamma \\ \gamma & \gamma & 1 & \cdots & \gamma \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma & \gamma & \gamma & \cdots & 1 \end{bmatrix} \quad (21)$$

for $\gamma \in [0, 1]$. Since it can be written as

$$G_n(\gamma) = \gamma \mathbf{1}\mathbf{1}^T + (1 - \gamma)\mathbb{I}, \quad (22)$$

where $\mathbf{1}$ is the all-ones column vector and $\mathbb{I}$ is the $n \times n$ identity matrix, we know that it is always positive semidefinite for this range of γ . Therefore, it is the Gram matrix of some collection of vectors and since the diagonal entries of $G_n(\gamma)$ are all equal to 1, those vectors can be interpreted as pure states.

We state a useful lemma about equiangular states.

Lemma 11. For $n \geq 2, k \in \{1, 2, \dots, n\}$, we have that $G_n(\gamma)$ is k -incoherent when $0 \leq \gamma \leq \frac{k-1}{n-1}$.

Proof. This proof is similar to one in [JRS25] which proved it for the case of $k = n - 1$, corresponding to antidistinguishability. The case of $k = n$ is trivial as every $n \times n$ positive semidefinite matrix is n -incoherent. The case of $k = 1$ is also trivial as the condition enforces $\gamma = 0$, i.e., a diagonal matrix, which is always 1-incoherent. The rest of the proof proves the case when $k \in \{2, \dots, n - 1\}$ and, by extension, $n \geq 3$.

Set $x = (k - 1)/(n - 1)$ for notational clarity and, for each $I \subseteq \{1, 2, \dots, n\}$ with $|I| = n - k$, define

$$G_I := \frac{1}{\binom{n-2}{n-k}} \left((x - \gamma)(\mathbb{1} - \sum_{i \in I} \mathbf{e}_i \mathbf{e}_i^\top) + \gamma \left(\mathbf{1} - \sum_{i \in I} \mathbf{e}_i \right) \left(\mathbf{1} - \sum_{i \in I} \mathbf{e}_i \right)^\top \right). \quad (23)$$

It is clear that G_I is positive semidefinite (since $\gamma \leq (k - 1)/(n - 1) = x$) and is equal to 0 outside of the $k \times k$ principal submatrix that avoids rows and columns with indices in I . Furthermore, direct calculation shows that

$$\begin{aligned} \sum_{\substack{I \subseteq \{1, 2, \dots, n\} \\ |I| = n-k}} G_I &= \frac{1}{\binom{n-2}{n-k}} \sum_{\substack{I \subseteq \{1, 2, \dots, n\} \\ |I| = n-k}} \left((x - \gamma)(\mathbb{1} - \sum_{i \in I} \mathbf{e}_i \mathbf{e}_i^\top) + \gamma \left(\mathbf{1} - \sum_{i \in I} \mathbf{e}_i \right) \left(\mathbf{1} - \sum_{i \in I} \mathbf{e}_i \right)^\top \right) \\ &= \frac{1}{\binom{n-2}{n-k}} \left((x - \gamma) \binom{n-1}{n-k} \mathbb{1} + \gamma \binom{n-1}{n-k} (x \mathbf{1} \mathbf{1}^\top + (1 - x) \mathbb{1}) \right) \\ &= \frac{x \binom{n-1}{n-k}}{\binom{n-2}{n-k}} ((1 - \gamma) \mathbb{1} + \gamma \mathbf{1} \mathbf{1}^\top) \\ &= G_n(\gamma), \end{aligned} \quad (24)$$

so it follows that $G_n(\gamma)$ is k -incoherent by definition. $\square$

By Proposition 10, we know that the equiangular states with $0 \leq \gamma \leq \frac{k-1}{n-1}$ must be k -learnable. On the other hand, we have

$$\sum_{i,j=1}^n |\langle \psi_i | \psi_j \rangle| = \sum_{i,j=1}^n |G_{i,j}| = n + n(n-1)\gamma. \quad (25)$$

Thus by Theorem 3, if $\gamma > \frac{k-1}{n-1}$, then it cannot be k -learnable. Combining, we have the following corollary.

Corollary 12. *Let $0 \leq \gamma \leq 1$, $n \geq 2$, $k \in \{1, 2, \dots, n\}$, and let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$ be such that $\langle \psi_i | \psi_j \rangle = \gamma$ whenever $i \neq j$. In other words, S has Gram matrix given by Equation (21). Then S is k -learnable if and only if*

$$\gamma \leq \frac{k-1}{n-1}. \quad (26)$$

By Corollary 12, setting $\gamma = \frac{k-1}{n-1}$ yields a set of n pure states that are k -learnable, proving that both bounds in Theorem 3 are sharp.

3 Sharpness of the Frobenius norm bound of Inequality (3)

We now show that Inequality (3) of Theorem 2 is tight in the sense that the right-hand side of the inequality cannot be increased at all. To this end, note that

$$\sum_{i,j=1}^n |\langle \psi_i | \psi_j \rangle|^2 = \|G\|_{\text{F}}^2,$$

where G is the Gram matrix of $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle$ and $\|\cdot\|_F$ denotes the Frobenius norm. We prove tightness by demonstrating how to construct a Gram matrix that is not k -incoherent (and thus corresponds to a set of states that is not k -learnable) with any Frobenius norm that is slightly larger than $n/\sqrt{n-k+1}$.

Theorem 13. *Let $1 < k < n$ be integers, let $r := n - k + 1$, and let $\omega := e^{\frac{2\pi i}{n}}$. For each $j \in \{1, 2, \dots, r\}$, let*

$$\mathbf{v}_j = \left(1, \omega^{j-1}, \omega^{2(j-1)}, \dots, \omega^{(n-1)(j-1)}\right) \in \mathbb{C}^n. \quad (27)$$

Consider the Gram matrix

$$G := \frac{1}{r} \sum_{j=1}^r \mathbf{v}_j \mathbf{v}_j^*. \quad (28)$$

Then G has $\|G\|_F = n/\sqrt{r}$ and is k -incoherent but not $(k-1)$ -incoherent. Furthermore, for any $\varepsilon \in (0, 1)$, the Gram matrix

$$G_\varepsilon := \frac{1}{r} \left((1+\varepsilon) \mathbf{v}_1 \mathbf{v}_1^* + (1-\varepsilon) \mathbf{v}_r \mathbf{v}_r^* + \sum_{i=2}^{r-1} \mathbf{v}_i \mathbf{v}_i^* \right) \quad (29)$$

is not k -incoherent.

In particular, since

$$\lim_{\varepsilon \rightarrow 0^+} G_\varepsilon = G \quad \text{and} \quad \lim_{\varepsilon \rightarrow 0^+} \|G_\varepsilon\|_F = \lim_{\varepsilon \rightarrow 0^+} \left(\frac{n}{\sqrt{r}} \sqrt{1 + \frac{2\varepsilon^2}{r}} \right) = \frac{n}{\sqrt{r}}, \quad (30)$$

Theorem 13 shows that Inequality (3) is tight: for any Frobenius norm slightly above $n/\sqrt{r}$, we can choose $\varepsilon > 0$ so as to create a Gram matrix G_ε with that Frobenius norm that is not k -incoherent, so (by Proposition 10) the associated set of states is not k -learnable.

Proof of Theorem 13. The fact that G and G_ε are Gram matrices is straightforward to verify. Since the vectors $\mathbf{v}_j$ are mutually orthogonal, the Gram matrix G is proportional to a rank- r orthogonal projection, and it is easily verified that $\|G\|_F = n/\sqrt{r}$. By Theorem 2, G is k -incoherent.

Let $\mathbf{y} \in \text{im}(G)$ (the image of G) be a vector with at most $n - r = k - 1$ nonzero entries, and let $j_1, \dots, j_r$ be indices for which $y_{j_1} = \dots = y_{j_r} = 0$. Since $\mathbf{v}_1, \dots, \mathbf{v}_r$ form a basis of $\text{im}(G)$, there exist scalars $\{\alpha_j\} \subset \mathbb{C}$ such that $\mathbf{y} = \sum_{j=1}^r \alpha_j \mathbf{v}_j$. It follows that

$$\sum_{j=1}^r \alpha_j \omega^{(j_\ell-1)(j-1)} = 0 \quad (31)$$

for all $\ell \in \{1, 2, \dots, r\}$. If we define the polynomial $p(x) = \sum_{j=1}^r \alpha_j x^{j-1}$ then Equation (31) says exactly that

$$p(\omega^{j_1-1}) = \dots = p(\omega^{j_r-1}) = 0. \quad (32)$$

Since p is a polynomial of degree at most $r - 1$ and has r roots, it must be the case that $p(x) = 0$ for all $x \in \mathbb{C}$. It follows that $\alpha_1 = \dots = \alpha_r = 0$, so $\mathbf{y} = \mathbf{0}$. We have thus shown that there are no nonzero vectors in $\text{im}(G)$ with at most $k - 1$ nonzero entries, so G is not $(k - 1)$ -incoherent.

It remains only to prove that G_ε is not k -incoherent. Similar to before, let $\mathbf{y} \in \text{im}(G_\varepsilon)$ have at most $k = n - r + 1$ nonzero entries, and let $j_1, \dots, j_{r-1}$ be indices for which $y_{j_1} = \dots = y_{j_{r-1}} = 0$. Then there exist scalars $\{\alpha_j\} \subset \mathbb{C}$ (in particular, $\alpha_j = \mathbf{v}_j^* \mathbf{y} / n$) such that $\mathbf{y} = \sum_{j=1}^r \alpha_j \mathbf{v}_j$, so

$$\sum_{j=1}^r \alpha_j \omega^{(j_\ell-1)(j-1)} = 0 \quad (33)$$

for all $\ell \in \{1, 2, \dots, r-1\}$. If we define the polynomial $p(x) = \sum_{j=1}^r \alpha_j x^{j-1}$ then Equation (33) says exactly that

$$p(\omega^{j_1-1}) = \dots = p(\omega^{j_{r-1}-1}) = 0. \quad (34)$$

Since p is a polynomial of degree at most $r-1$, and we have found $r-1$ roots of it, it must have the form $p(x) = \alpha_r(x - \omega^{j_1-1}) \dots (x - \omega^{j_{r-1}-1})$. It follows that $\alpha_1 = \pm \alpha_r(\omega^{j_1-1} \dots \omega^{j_{r-1}-1})$, so $|\alpha_1| = |\alpha_r|$, which implies

$$|\mathbf{v}_1^* \mathbf{y}| = |\mathbf{v}_r^* \mathbf{y}|. \quad (35)$$

Now suppose (for the sake of establishing a contradiction) that G_ε is k -incoherent, so that it has a decomposition of the form $G_\varepsilon = \sum_i \mathbf{y}_i \mathbf{y}_i^*$, where each $\mathbf{y}_i$ has at most k nonzero entries. Since each $\mathbf{v}_j$ is an eigenvector of G_ε , it follows that the associated eigenvalue λ_j is given by

$$\lambda_j = \frac{1}{n} \mathbf{v}_j^* G_\varepsilon \mathbf{v}_j = \frac{1}{n} \sum_i |\mathbf{v}_j^* \mathbf{y}_i|^2. \quad (36)$$

Notice that the form of G_ε given in the statement of the theorem tells us that if we sort the eigenvalues non-increasingly then we have $\lambda_1 = n(1 + \varepsilon)/r$, $\lambda_j = n/r$ for $j \in \{2, 3, \dots, r-1\}$, and $\lambda_r = n(1 - \varepsilon)/r$. Since $\mathbf{y}_i \in \text{im}(G_\varepsilon)$ for all i , Equation (35) tells us that $|\mathbf{v}_1^* \mathbf{y}_i| = |\mathbf{v}_r^* \mathbf{y}_i|$ for all i . But Equation (36) then implies

$$\frac{n}{r}(1 + \varepsilon) = \lambda_1 = \lambda_r = \frac{n}{r}(1 - \varepsilon), \quad (37)$$

which is a contradiction that completes the proof. $\square$

Remark 14. The proof of Theorem 13 provides a way of constructing a Gram matrix that is not k -incoherent and whose spectrum is arbitrarily close to

$$\frac{n}{n-k+1}, \frac{n}{n-k+1}, \dots, \frac{n}{n-k+1}, 0, 0, \dots, 0, \quad (38)$$

where the eigenvalue $n/(n-k+1)$ occurs $n-k+1$ times. In the terminology of [JMPP22], the same argument shows that (38) is the only spectrum belonging to an absolutely k -incoherent matrix of trace n and rank at most $n-k+1$. Theorem 13 can thus be regarded as a strengthening of [JMPP22, Theorem 4], which showed the weaker statement that no matrix with rank strictly less than $n-k+1$ is absolutely k -incoherent.

4 Packing bounds for k -learnability

When $k = 1$, k -learnability of a multiset of states is equivalent to those states being mutually orthogonal. As a result, the number of states n in a 1-learnable set cannot exceed the dimension d . The following theorem generalizes this observation to higher values of k , establishing bounds on the number of states in a k -learnable multiset in terms of the dimension.

Theorem 15. Let $\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be a multiset of pure states that is k -learnable.

(1) It holds that $n \leq dk$.

(2) If the vectors are pairwise non-parallel, i.e., $|\psi_i\rangle \notin \text{span}\{|\psi_j\rangle\}$ for all $i \neq j$, then $2n \leq d(k+1)$.

Proof. We may assume that $\dim(\text{span}\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}) = d$, as this can only make the desired inequality tighter. Let G be the Gram matrix of this multiset, and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{C}^n$ be vectors with at most k nonzero entries such that

$$G = \sum_{j=1}^m \mathbf{v}_j \mathbf{v}_j^*. \quad (39)$$

It follows that $\text{im}(G)$ (the image of G) is spanned by $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$. After permuting terms, we may assume that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d\}$ forms a basis for $\text{im}(G)$. Let $A = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d] \in \mathbb{C}^{n \times d}$ and $Y = [|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle]^* \in \mathbb{C}^{n \times d}$. Since $\text{im}(A) = \text{im}(G) = \text{im}(Y)$, there exists an invertible matrix $C \in \mathbb{C}^{d \times d}$ for which $A = YC$.

We first prove the bound $n \leq dk$. Since the vectors $|\psi_j\rangle$ are nonzero, it follows that the rows of A are nonzero. Let s denote the total number of nonzero entries of A . Since every row is nonzero, we have $s \geq n$. On the other hand, A has d columns and each column has support size at most k , so $s \leq kd$. Combining the two bounds gives $n \leq kd$.

Now we prove the bound $2n \leq d(k+1)$ under the assumption that the vectors are pairwise non-parallel. Note that the rows of Y are pairwise non-parallel, as these are the vectors $|\psi_j\rangle^*$. Since C is invertible, the rows of A are also pairwise non-parallel. Let t be the number of rows of A having exactly one nonzero entry. Any such row is a nonzero scalar multiple of one of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d$. Since the rows are pairwise non-parallel, there can be at most one such row for each coordinate, so $t \leq d$. Let s denote the total number of nonzero entries of A . Since every row is nonzero, and every row that is not a scalar multiple of an $\mathbf{e}_j$ has support size at least 2, we have

$$s \geq t + 2(n - t) = 2n - t \geq 2n - d. \quad (40)$$

On the other hand, A has d columns and each column has support size at most k , so $s \leq kd$. Combining the two bounds gives $2n - d \leq kd$, hence $2n \leq d(k+1)$, which completes the proof. $\square$

Next, we show that the bounds of Theorem 15 are tight: for any tuple of integers (k, d, n) satisfying the inequalities specified by that theorem, there exists a k -learnable multiset of n pure states in $\mathbb{C}^d$.

Theorem 16. Let $k, d \leq n$ be positive integers.

(1) If $n \leq dk$, then there exists a multiset of pure states $\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ that is k -learnable.

(2) If $2n \leq d(k+1)$, then there exists a k -learnable multiset of pure states $\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ that is pairwise non-parallel, i.e., $|\psi_i\rangle \notin \text{span}\{|\psi_j\rangle\}$ for all $i \neq j$.

Proof. We begin by proving the first statement. Write $n = qk + s$ for a positive integer q and $0 \leq s < k$. Consider the following n unit vectors:

$$\underbrace{\mathbf{e}_1, \dots, \mathbf{e}_1}_{k \text{ times}}, \underbrace{\mathbf{e}_2, \dots, \mathbf{e}_2}_{k \text{ times}}, \dots, \underbrace{\mathbf{e}_q, \dots, \mathbf{e}_q}_{k \text{ times}}, \underbrace{\mathbf{e}_{q+1}, \dots, \mathbf{e}_{q+1}}_s. \quad (41)$$

These vectors span a space of dimension at most d , and are clearly k -learnable.

Now we prove the second statement. Let $m = n - d$, so $2m = 2n - 2d \leq d(k - 1)$. Let H be a multigraph on vertex set $\{1, 2, \dots, d\}$ with m edges and maximum vertex degree at most $k - 1$. To construct H , start with the edgeless graph on d vertices and at each step add an edge between any two vertices whose degrees are at most the degrees of all of the remaining $d - 2$ vertices. One can easily check that the difference between the maximum and minimum vertex degree at any step is at most 1. Furthermore, the inequality $2m \leq d(k - 1)$ guarantees that if m edges are added in this way, no vertex will have degree greater than $k - 1$.

Let $A \in \mathbb{C}^{n \times d}$ be defined as follows. First include the d standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d$ as rows. Next, list the edges of H as $\{i_t, j_t\}$ for $t \in \{1, 2, \dots, m\}$. Choose phases $e^{i\phi_t}$ (distinct if parallel edges occur) and define additional rows of A to be

$$\mathbf{w}_t := \frac{1}{\sqrt{2}} \left(\mathbf{e}_{i_t} + e^{i\phi_t} \mathbf{e}_{j_t} \right) \quad \text{for } t \in \{1, 2, \dots, m\}. \quad (42)$$

It follows that A has $d + m = n$ rows, and every row has norm 1, so $G := AA^*$ is a Gram matrix of n unit vectors. The first d rows of A are $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d$, so $\text{rank}(A) = d$, which implies $\text{rank}(G) = \text{rank}(AA^*) = d$.

Let $\mathbf{v}_j$ denote the j -th column of A . Then

$$G = AA^* = \sum_{j=1}^d \mathbf{v}_j \mathbf{v}_j^*. \quad (43)$$

Notice that $\mathbf{v}_j$ is nonzero in its j -th entry (since that entry comes from the row $\mathbf{e}_j$ in A) and in the entries corresponding to edges incident to vertex j . Hence

$$|\text{supp}(\mathbf{v}_j)| = 1 + \deg_H(j) \leq k. \quad (44)$$

In particular, this implies that G is k -incoherent. Hence, the complex conjugates of the rows of A form a k -learnable multiset of pure states, which completes the proof. $\square$

5 Copy complexity bounds for k -learnability

The number of copies of an unknown quantum state required to perform a learning task is commonly called its *sample complexity* or *copy complexity*. It has been studied for problems including state tomography [HHJ⁺17, Mon17], PAC and shadow learning [Aar07, Aar18, HKP20], state certification [BOW19], and property testing [OW21]. Here we study an exact, zero-error version of copy complexity: given a finite multiset S , we ask for the minimum number of copies after which every measurement outcome identifies a list of at most k candidates containing the unknown state. Since

$$\langle \psi_i^{\otimes c} | \psi_j^{\otimes c} \rangle = \langle \psi_i | \psi_j \rangle^c,$$

taking additional copies suppresses nontrivial overlaps, and we determine when this makes the resulting states k -learnable.

Definition 17. Let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be a multiset of pure states. We say that S is (c, k) -learnable if the multiset $S^{(c)} := \{|\psi_1\rangle^{\otimes c}, \dots, |\psi_n\rangle^{\otimes c}\}$ is k -learnable.

We define the copy complexity of k -learning for S , denoted $c_k(S)$, to be the minimum c for which S is (c, k) -learnable. If no finite c exists, then we define $c_k(S) = \infty$.

5.1 Copy complexity bounds from maximum pairwise overlap

Our main theorem provides a generic copy complexity bound for 2-learning an arbitrary multiset in terms of the overlap structure of the multiset of states.

Corollary 18. *Let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be a multiset of $n \geq 3$ pure states with maximum pairwise overlap $0 < \delta := \max_{i \neq j} |\langle \psi_i | \psi_j \rangle|^2 < 1$. Then S is $(c, 2)$ -learnable for*

$$c = \left\lceil \frac{2 \log(n-1)}{\log(1/\delta)} \right\rceil. \quad (45)$$

Proof. The squared Frobenius norm of the Gram matrix $G^{(c)}$ of $S^{(c)} = \{|\psi_i\rangle^{\otimes c}\}$ satisfies

$$\|G^{(c)}\|_F^2 = \sum_{i,j=1}^n |\langle \psi_i | \psi_j \rangle|^{2c} = n + \sum_{i \neq j} |\langle \psi_i | \psi_j \rangle|^{2c} \leq n + n(n-1)\delta^c. \quad (46)$$

By Theorem 2 with $k = 2$, S is $(c, 2)$ -learnable whenever $\|G^{(c)}\|_F^2 \leq n^2/(n-1)$. This holds when

$$n(n-1)\delta^c \leq \frac{n}{n-1}, \quad (47)$$

i.e., $\delta^c \leq (n-1)^{-2}$. Taking logarithms gives $c \geq 2 \log(n-1) / \log(1/\delta)$. $\square$

5.2 SIC-POVMs and mutually unbiased bases

In this section we determine copy complexity bounds for k -learnability of SIC-POVMs and mutually unbiased bases. We use a straightforward result on (c, k) -learnability of complex projective c -designs, which we soon state.

We say that a multiset of states $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ forms a (complex projective) c -design if

$$\sum_{i=1}^n |\psi_i\rangle \langle \psi_i|^{\otimes c} = \frac{n}{\binom{d+c-1}{c}} \Pi_{d,c}, \quad (48)$$

where $\Pi_{d,c}$ is the projection to $S^c(\mathbb{C}^d) \subseteq (\mathbb{C}^d)^{\otimes c}$. Here, $S^c(\mathbb{C}^d)$ is the symmetric subspace of vectors $\mathbf{v} \in (\mathbb{C}^d)^{\otimes c}$ for which $v_{i_1, \dots, i_c} = v_{i_{\sigma(1)}, \dots, i_{\sigma(c)}}$ for all permutations $\sigma \in \mathfrak{S}_c$.

Corollary 19. *Let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be a multiset of pure states that forms a c -design. Then S is (c, k) -learnable for $k = n - \binom{d+c-1}{c} + 1$.*

Proof. Let $A = [|\psi_1\rangle^{\otimes c}, \dots, |\psi_n\rangle^{\otimes c}]$ be the $d^c \times n$ matrix with columns $|\psi_i\rangle^{\otimes c}$. By the definition of c -design, it holds that

$$AA^* = \sum_{i=1}^n |\psi_i\rangle \langle \psi_i|^{\otimes c} = \frac{n}{\binom{d+c-1}{c}} \Pi_{d,c}. \quad (49)$$

Note that $G = A^*A$ has the same nonzero eigenvalues as AA^* , which are $n \binom{d+c-1}{c}^{-1}$ with multiplicity $\binom{d+c-1}{c}$. It follows that

$$\|G\|_F = \frac{n}{\sqrt{\binom{d+c-1}{c}}}. \quad (50)$$

The corollary now follows immediately from Theorem 2. $\square$

5.2.1 SIC-POVMs

A *symmetric informationally complete POVM* (SIC-POVM) is a set $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_{d^2}\rangle\} \subset \mathbb{C}^d$ for which $|\langle\psi_i|\psi_j\rangle|^2 = 1/(d+1)$ for all $i \neq j$. We first use Proposition 4 to determine the copy complexity of 2-learning a SIC-POVM.

Corollary 20. *Let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_{d^2}\rangle\} \subset \mathbb{C}^d$ be a SIC-POVM. Then*

$$c_2(S) = \begin{cases} 2, & d = 2, \\ 3, & d = 3, \\ 4, & d \geq 4. \end{cases} \quad (51)$$

Proof. For a SIC-POVM, $|\langle\psi_i|\psi_j\rangle|^2 = 1/(d+1)$ whenever $i \neq j$. Hence the Gram matrix of $S^{(c)} = \{|\psi_1\rangle^{\otimes c}, \dots, |\psi_{d^2}\rangle^{\otimes c}\}$ satisfies

$$R := \sum_{j=1}^{d^2} |\langle\psi_i|\psi_j\rangle|^c = 1 + \frac{d^2 - 1}{\sqrt{(d+1)^c}} \quad (52)$$

for all $i \in \{1, 2, \dots, d^2\}$. By Proposition 4, S is $(c, 2)$ -learnable if and only if $R \leq 2$, or equivalently,

$$c \geq \frac{2 \log(d^2 - 1)}{\log(d+1)}. \quad (53)$$

Therefore

$$c_2(S) = \left\lceil \frac{2 \log(d^2 - 1)}{\log(d+1)} \right\rceil. \quad (54)$$

One can easily verify that this simplifies to the expression in the corollary statement. $\square$

The above corollary shows that 4 copies always suffice to 2-learn a SIC-POVM. We now use Theorem 2 to provide a bound on the k -learnability of a SIC-POVM if we are given fewer than 4 copies.

Corollary 21. *Let $S \subset \mathbb{C}^d$ be a SIC-POVM. Then S is*

$$k\text{-learnable} \quad \text{for} \quad k = d^2 - d + 1, \quad (55)$$

$$(2, k)\text{-learnable} \quad \text{for} \quad k = \binom{d}{2} + 1, \quad (56)$$

$$(3, k)\text{-learnable} \quad \text{for} \quad k = \left\lceil \frac{d^2 + 3}{d + 3} \right\rceil. \quad (57)$$

Proof. It is known that SIC-POVMs form 1-designs and 2-designs. Hence, S is k -learnable for $k = d^2 - d + 1$, and $(2, k)$ -learnable for $k = \binom{d}{2} + 1$ by Corollary 19. In general, for a SIC-POVM S , the Gram matrix $G^{(c)}$ of $S^{(c)}$ satisfies

$$\|G^{(c)}\|_F^2 = d^2 + \frac{d^4 - d^2}{(d+1)^c}. \quad (58)$$

By Theorem 2, S is (c, k) -learnable if

$$d^2 + \frac{d^4 - d^2}{(d+1)^c} \leq \frac{d^4}{d^2 - k + 1}. \quad (59)$$

Rearranging, we have that S is (c, k) -learnable for

$$k = 1 + \left\lceil \frac{d^2(d^2 - 1)}{(d + 1)^c + d^2 - 1} \right\rceil. \quad (60)$$

For $c = 3$, this gives

$$k = \left\lceil \frac{d^2 + 3}{d + 3} \right\rceil. \quad (61)$$

This completes the proof. $\square$

5.2.2 Mutually unbiased bases

A collection of *mutually unbiased bases* (MUBs) is a set

$$S = \{|\psi_{i,j}\rangle : i \in \{1, 2, \dots, d\}, j \in \{1, 2, \dots, s\}\} \subset \mathbb{C}^d \quad (62)$$

for some integer $s \geq 2$ such that $\{|\psi_{i,j}\rangle : i \in \{1, 2, \dots, d\}\}$ is an orthonormal basis of $\mathbb{C}^d$ for each j , and $|\langle \psi_{i,j} | \psi_{i',j'} \rangle|^2 = 1/d$ for all i, i' and $j \neq j'$. It is known that $s \leq d + 1$ must hold for any collection of MUBs.

We first use Proposition 4 to determine the exact copy complexity of 2-learning a collection of MUBs.

Corollary 22. *Let $S = \{|\psi_{i,j}\rangle : i \in \{1, 2, \dots, d\}, j \in \{1, 2, \dots, s\}\} \subset \mathbb{C}^d$ be a collection of $s \geq 2$ mutually unbiased bases. Then*

$$c_2(S) = \lceil 2 + 2 \log_d(s - 1) \rceil. \quad (63)$$

In particular, $c_2(S) \leq 4$, with equality for a complete collection of $d + 1$ mutually unbiased bases.

Proof. Fix $|\psi_{i,j}\rangle \in S$. The states in the same basis are orthogonal, while for every $j' \neq j$ and every $i' \in \{1, 2, \dots, d\}$, we have

$$|\langle \psi_{i,j} | \psi_{i',j'} \rangle| = \frac{1}{\sqrt{d}}. \quad (64)$$

It follows that the sum of the absolute values of any row in the Gram matrix of

$$S^{(c)} = \{|\psi_{i,j}\rangle^{\otimes c} : i \in \{1, 2, \dots, d\}, j \in \{1, 2, \dots, s\}\} \quad (65)$$

is given by $R = 1 + (s - 1)d^{1-c/2}$. By Proposition 4, S is $(c, 2)$ -learnable if and only if $R \leq 2$, or equivalently if and only if

$$c \geq 2 + 2 \log_d(s - 1). \quad (66)$$

This proves the general expression for $c_2(S)$.

Finally, since $s \leq d + 1$ for any collection of mutually unbiased bases, $\log_d(s - 1) \leq 1$, so $c_2(S) \leq 4$. If $s = d + 1$, then $c_2(S) = \lceil 2 + 2 \log_d d \rceil = 4$, so the bound is attained for a maximal collection of MUBs. $\square$

The previous corollary showed that 4 copies of a collection of MUBs always suffice to 2-learn them. The following corollary tells us what (weaker) type of learning is possible if we have fewer than 4 copies.

Corollary 23. Let $S = \{|\psi_{i,j}\rangle : i \in \{1, 2, \dots, d\}, j \in \{1, 2, \dots, s\}\} \subset \mathbb{C}^d$ be a collection of $s \geq 2$ mutually unbiased bases. Then S is

$$k\text{-learnable} \quad \text{for} \quad k = sd - d + 1, \quad (67)$$

$$(2, k)\text{-learnable} \quad \text{for} \quad k = 1 + \left\lceil \frac{s(s-1)d}{d+s-1} \right\rceil, \quad (68)$$

$$(3, k)\text{-learnable} \quad \text{for} \quad k = 1 + \left\lceil \frac{s(s-1)d}{d^2+s-1} \right\rceil. \quad (69)$$

Proof. Let $G^{(c)}$ be the Gram matrix of $S^{(c)}$. By direct calculation,

$$\|G^{(c)}\|_F^2 = sd + \frac{s(s-1)d^2}{d^c}. \quad (70)$$

By Theorem 2, S is (c, k) -learnable if

$$\|G^{(c)}\|_F^2 \leq \frac{(sd)^2}{sd - k + 1}. \quad (71)$$

Rearranging, it is sufficient to set

$$k = 1 + \left\lceil \frac{s(s-1)d}{d^{c-1} + s - 1} \right\rceil. \quad (72)$$

This reproduces the above expressions, and completes the proof. $\square$

5.3 Stabilizer states

In the following, for a prime number d we let $\text{Stab}_{d,m} \subset (\mathbb{C}^d)^{\otimes m}$ be the set of m -qudit stabilizer states, and let $s_{d,m} := |\text{Stab}_{d,m}| = d^m \prod_{j=1}^m (d^j + 1)$ (see [Gro06, Corollary 21] or [KG15, Proposition 2]). For an integer $c \geq 1$, let $G_{d,m}^{(c)}$ be the Gram matrix of the set

$$\{|\psi\rangle^{\otimes c} : |\psi\rangle \in \text{Stab}_{d,m}\}. \quad (73)$$

Proposition 24. For every prime number d and all positive integers m and c it holds that

$$\|G_{d,m}^{(c)}\|_F^2 = s_{d,m} \prod_{j=0}^{m-1} (1 + d^{j-c+2}). \quad (74)$$

The proof of Proposition 24 is in Appendix A.4.

5.3.1 2-learning stabilizer states

We now apply Proposition 4 to stabilizer states. Fix a prime d , and let

$$\text{Stab}_{d,m} \subset (\mathbb{C}^d)^{\otimes m} \quad (75)$$

denote the set of m -qudit stabilizer states. The Clifford group acts transitively on $\text{Stab}_{d,m}$, so for a fixed $|\psi\rangle \in \text{Stab}_{d,m}$ the quantity

$$\Phi_c(d, m) := \sum_{\phi \in \text{Stab}_{d,m}} |\langle \psi | \phi \rangle|^c \quad (76)$$

is independent of $|\psi\rangle$. We first determine this quantity exactly.

Proposition 25. For every prime d , every positive integer m , and every positive integer c ,

$$\Phi_c(d, m) = \prod_{j=0}^{m-1} \left(1 + d^{j+2-c/2}\right). \quad (77)$$

Proof. For a fixed stabilizer state $|\psi\rangle$, let a_r denote the number of stabilizer states $|\phi\rangle$ satisfying

$$|\langle\psi|\phi\rangle|^2 = d^{-r}, \quad r \in \{0, 1, \dots, m\}. \quad (78)$$

The possible nonzero overlaps between stabilizer states have precisely this form [OHSG24, Equation (41)] (see also [GMC17, Theorem 11]). Hence, for every positive integer t ,

$$\sum_{r=0}^m a_r d^{-rt} = \sum_{\phi \in \text{Stab}_{d,m}} |\langle\psi|\phi\rangle|^{2t}. \quad (79)$$

By Proposition 24 and Clifford transitivity,

$$\sum_{\phi \in \text{Stab}_{d,m}} |\langle\psi|\phi\rangle|^{2t} = \prod_{j=0}^{m-1} \left(1 + d^{j-t+2}\right). \quad (80)$$

Therefore

$$\sum_{r=0}^m a_r x^r = \prod_{j=0}^{m-1} (1 + d^{j+2} x) \quad (81)$$

for the infinitely many values $x = d^{-t}$ ($t \geq 1$ is a positive integer). Since both sides are polynomials of degree at most m , this is a polynomial identity. Setting $x = d^{-c/2}$ gives

$$\Phi_c(d, m) = \sum_{r=0}^m a_r d^{-rc/2} = \prod_{j=0}^{m-1} \left(1 + d^{j+2-c/2}\right), \quad (82)$$

as claimed. $\square$

Combining Propositions 4 and 25 gives an exact criterion for 2-learning stabilizer states.

Theorem 26. Let d be prime. Then $\text{Stab}_{d,m}$ is $(c, 2)$ -learnable if and only if

$$\prod_{j=0}^{m-1} \left(1 + d^{j+2-c/2}\right) \leq 2. \quad (83)$$

This criterion allows us to determine exactly the copy complexity of 2-learning stabilizer states.

Corollary 27. Let d be prime. The copy complexity of 2-learning the set of m -qudit stabilizer states is

$$c_2(\text{Stab}_{d,m}) = \begin{cases} 4, & m = 1, \\ 8, & d = 2, m = 2, \\ 2m + 5, & d = 2, m \geq 3, \\ 7, & d = 3, m = 2, \\ 2m + 4, & d = 3, m \geq 3, \\ 2m + 3, & d \geq 5, m \geq 2. \end{cases} \quad (84)$$

In particular, $c_2(\text{Stab}_{d,m}) = 2m + \Theta(1)$.

The proof of Corollary 27 can be found in Appendix A.5.

Remark 28. Any 2-learning measurement for a set S can be converted into a bounded-error identification procedure. First use $c_2(S)$ copies to obtain a certified pair $\{|\alpha\rangle, |\beta\rangle\}$ containing the unknown state. Then use t additional copies and project onto $|\alpha\rangle^{\otimes t}$. If the unknown state is $|\alpha\rangle$, the test never errs; if it is $|\beta\rangle$, the error probability is

$$|\langle\alpha|\beta\rangle|^{2t}. \quad (85)$$

For two distinct qudit stabilizer states,

$$|\langle\alpha|\beta\rangle|^2 \leq \frac{1}{d}. \quad (86)$$

Thus $t = m$ additional copies reduce the error probability to at most d^{-m} , giving an information-theoretic identification procedure using $c_2(S) + m = 3m + O(1)$ copies in total. This reproduces the information-theoretic $O(m)$ upper bound proven in [Mon17, AG08, ADIS25, CCLW26], with the caveat that our result applies only for prime-dimensional qudits. A drawback to our result is that the 2-learning measurement may not be efficiently implementable. An advantage is that the 2-learning approach additionally yields a distinguished pair $\{|\alpha\rangle, |\beta\rangle\}$ of stabilizer states guaranteed to contain the unknown state. Note that $\Omega(m)$ copies are necessary: There are $d^{\Theta(m^2)}$ stabilizer states on m d -dimensional qudits, so identifying an unknown stabilizer state to constant error probability requires $\Omega(m)$ copies by Holevo's theorem.

5.3.2 k -learning stabilizer states

Beyond the case $k = 2$, we can use the above results to prove a general relationship between c and k for stabilizer states.

Corollary 29. Let d be prime and let m and c be positive integers. The set of stabilizer states $\text{Stab}_{d,m} \subset (\mathbb{C}^d)^{\otimes m}$ is (c, k) -learnable for

$$k = \left\lceil s_{d,m} \left(1 - \frac{1}{\prod_{j=0}^{m-1} (1 + d^{j-c+2})} \right) \right\rceil + 1. \quad (87)$$

Proof. By Theorem 2 and Proposition 24, the set of stabilizer states is (c, k) -learnable as long as

$$s_{d,m} \prod_{j=0}^{m-1} (1 + d^{j-c+2}) \leq \frac{s_{d,m}^2}{s_{d,m} - k + 1}. \quad (88)$$

Solving for k produces the quantity in Equation (87). $\square$

6 Zero-error quantum state mutation detection problems

We now apply Theorems 2 and 3 to families of product states that arise in anomaly and change-point detection. These families of pure states have a significant amount of symmetry, which allows us to obtain explicit zero-error k -learning criteria.

6.1 Single anomaly detection

Let $|\phi\rangle$ and $|\psi\rangle$ be pure states, and suppose that exactly one of n systems is prepared in the anomalous state $|\phi\rangle$, while all remaining systems are prepared in the reference state $|\psi\rangle$. If we do not

know which of the n systems is prepared in the anomalous state, then we know that the global state is of the form

$$|\psi\rangle^{\otimes(i-1)} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes(n-i)} \quad (89)$$

for some $i \in \{1, 2, \dots, n\}$. Determining the index i exactly can only be done if $\langle\phi|\psi\rangle = 0$. However, we can narrow i down to one of at most k possible values as long as $|\langle\phi|\psi\rangle|$ is small enough:

Theorem 30. *Let $n \geq 2$ be an integer and let $k \in \{1, 2, \dots, n\}$. The set*

$$\left\{ |\psi\rangle^{\otimes(i-1)} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes(n-i)} : i \in \{1, 2, \dots, n\} \right\} \quad (90)$$

is k -learnable if and only if

$$|\langle\phi|\psi\rangle| \leq \sqrt{\frac{k-1}{n-1}}. \quad (91)$$

Proof. For $i \neq j$, the two states $|\psi\rangle^{\otimes(i-1)} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes(n-i)}$ and $|\psi\rangle^{\otimes(j-1)} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes(n-j)}$ differ only in tensor factors i and j . Consequently, their inner product with each other is

$$\langle\phi|\psi\rangle \cdot \langle\psi|\phi\rangle = |\langle\phi|\psi\rangle|^2. \quad (92)$$

Since this is true for all $i \neq j$, the Gram matrix of this multiset of states is

$$G = (1 - \gamma)\mathbf{1}\mathbf{1}^\top + \gamma\mathbf{1}\mathbf{1}^\top, \quad (93)$$

where $\gamma = |\langle\phi|\psi\rangle|^2$ and $\mathbf{1}$ is the all-ones column vector. The result now follows immediately from Corollary 12. $\square$

In particular, Theorem 30 immediately implies that the smallest value of k for which

$$\left\{ |\psi\rangle^{\otimes(i-1)} \otimes |\phi\rangle \otimes |\psi\rangle^{\otimes(n-i)} : i \in \{1, 2, \dots, n\} \right\} \quad (94)$$

is k -learnable is $k = 1 + \lceil (n-1) |\langle\phi|\psi\rangle|^2 \rceil$.

6.2 Multi-anomaly detection

We next suppose that exactly $t \geq 1$ of the n systems are anomalous. Let $|\psi_0\rangle$ and $|\psi_1\rangle$ be pure states, and consider the global states of the form

$$|\psi_{i_1}\rangle \otimes \dots \otimes |\psi_{i_n}\rangle, \quad (95)$$

where $i_j \in \{0, 1\}$ for each $j \in \{1, 2, \dots, n\}$ and $\sum_{j=1}^n i_j = t$. We note that there are $\binom{n}{t}$ global states of this form.

Theorem 31. *Let $n \geq 2$ and $1 \leq t \leq n-1$ be integers, and define $\gamma := |\langle\psi_0|\psi_1\rangle|^2$. Then for each integer $1 \leq k \leq \binom{n}{t}$, the multiset of states*

$$S_t := \left\{ |\psi_{i_1}\rangle \otimes \dots \otimes |\psi_{i_n}\rangle : i_1, \dots, i_n \in \{0, 1\} \text{ and } \sum_{j=1}^n i_j = t \right\} \quad (96)$$

satisfies the following:

- (1) If $\sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^r > k$ then S_t is not k -learnable.
(2) If $(\binom{n}{t} - k + 1) \sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^{2r} \leq \binom{n}{t}$ then S_t is k -learnable.

Before proving this theorem, we note that if $r > t$ then we take the convention that $\binom{t}{r} = 0$. Also, in the $t = 1$ case we have

$$\sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^r = 1 + (n-1)\gamma = 1 + (n-1) |\langle \psi_0 | \psi_1 \rangle|^2, \quad (97)$$

so Theorem 31(1) says that if $1 + (n-1) |\langle \psi_0 | \psi_1 \rangle|^2 > k$ then S_1 is not k -learnable. This is equivalent to the $|\langle \phi | \psi \rangle| > \sqrt{(k-1)/(n-1)}$ condition from Theorem 30. However, Theorem 31(2) in the $t = 1$ case is weaker than the k -learnability condition from Theorem 30.

Proof of Theorem 31. Let G be the Gram matrix of S_t , indexed by the t -element subsets of $\{1, 2, \dots, n\}$. Let T and U be t -element subsets of $\{1, 2, \dots, n\}$ and set

$$r := |T \setminus U| = |U \setminus T| = t - |T \cap U|. \quad (98)$$

Every tensor factor indexed by $T \setminus U$ contributes $\langle \psi_0 | \psi_1 \rangle$ to the corresponding entry of G , and every tensor factor indexed by $U \setminus T$ similarly contributes $\langle \psi_1 | \psi_0 \rangle$. It follows that

$$G_{T,U} = (\langle \psi_0 | \psi_1 \rangle \langle \psi_1 | \psi_0 \rangle)^r = \gamma^r. \quad (99)$$

For each of the $\binom{n}{t}$ fixed t -element sets T , the number of t -element sets U satisfying $|T \setminus U| = r$ is $\binom{t}{r} \binom{n-t}{r}$, since we can choose r elements to remove from T and r elements to add back in from $\{1, 2, \dots, n\} \setminus T$. Since all entries of G are nonnegative, it follows that

$$\sum_{i,j} |G_{i,j}| = \binom{n}{t} \sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^r, \quad \text{and} \quad (100)$$

$$\|G\|_F^2 = \binom{n}{t} \sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^{2r}. \quad (101)$$

If

$$\sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^r > k, \quad (102)$$

then $\|\text{vec}(G)\|_1 > \binom{n}{t} k$, so Theorem 3 tells us that S_t is not k -learnable (recall that the number of states here is $\binom{n}{t}$), proving part (1) of this theorem. On the other hand, if

$$\left(\binom{n}{t} - k + 1 \right) \sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^{2r} \leq \binom{n}{t} \quad (103)$$

then

$$\|G\|_F^2 \leq \frac{\binom{n}{t}^2}{\binom{n}{t} - k + 1}, \quad (104)$$

so Theorem 2 tells us that S_t is k -learnable, proving part (2) of this theorem. $\square$

In particular, Theorem 31 tells us that the smallest k for which the multiset S_t from Equation (96) is k -learnable satisfies

$$\left\lceil \sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^r \right\rceil \leq k \leq \binom{n}{t} + 1 - \left\lfloor \frac{\binom{n}{t}}{\sum_{r=0}^n \binom{t}{r} \binom{n-t}{r} \gamma^{2r}} \right\rfloor. \quad (105)$$

6.3 Changepoint detection

Once again, let $|\phi\rangle$ and $|\psi\rangle$ be pure states and let n be a positive integer. Consider the states of the form $|\phi\rangle^{\otimes(i-1)} \otimes |\psi\rangle^{\otimes(n-i)}$ for some $i \in \{1, 2, \dots, n\}$ (i.e., the global states that locally equal $|\phi\rangle$ until they change to $|\psi\rangle$ at some point). There are n such states, each of which is $(n-1)$ -partite (if $i = 1$ then the state equals $|\psi\rangle$ on each party, and if $i = n$ then it equals $|\phi\rangle$ on each party).

Theorem 32. *Let $n \geq 2$ be an integer and let $\gamma := |\langle\phi|\psi\rangle| < 1$. Then for each integer $1 \leq k \leq n$, the multiset of states*

$$S := \left\{ |\phi\rangle^{\otimes(i-1)} \otimes |\psi\rangle^{\otimes(n-i)} : i \in \{1, 2, \dots, n\} \right\} \quad (106)$$

satisfies the following:

- (1) *If $2\gamma((n-1) - n\gamma + \gamma^n)/(1-\gamma)^2 > n(k-1)$ then S is not k -learnable.*
- (2) *If $2\gamma^2((n-1) - n\gamma^2 + \gamma^{2n})/(1-\gamma^2)^2 \leq n(k-1)/(n-k+1)$ then S is k -learnable.*

Proof. Let G be the Gram matrix of S . If $i < j$ then the i -th and j -th members of S differ precisely in the $j-i$ different tensor factors $i, i+1, \dots, j-1$, so

$$|G_{i,j}| = |\langle\phi|\psi\rangle|^{j-i} = \gamma^{j-i}. \quad (107)$$

If $i > j$ then we similarly have $|G_{i,j}| = \gamma^{i-j}$, so $|G_{i,j}| = \gamma^{|j-i|}$ for all i and j .

For each $r \in \{1, 2, \dots, n-1\}$, there are $2(n-r)$ ordered pairs $(i, j) \in \{1, 2, \dots, n\}^2$ with $|j-i| = r$. It follows that

$$\sum_{i,j} |G_{i,j}| = n + 2 \sum_{r=1}^{n-1} (n-r)\gamma^r = n + \frac{2\gamma((n-1) - n\gamma + \gamma^n)}{(1-\gamma)^2}, \quad (108)$$

$$\|G\|_F^2 = n + 2 \sum_{r=1}^{n-1} (n-r)\gamma^{2r} = n + \frac{2\gamma^2((n-1) - n\gamma^2 + \gamma^{2n})}{(1-\gamma^2)^2}, \quad (109)$$

where the closed-form expressions for the sum follow from the arithmetico-geometric formula (valid whenever $x \neq 1$):

$$\sum_{r=1}^{n-1} (n-r)x^r = \frac{x((n-1) - nx + x^n)}{(1-x)^2}. \quad (110)$$

Parts (1) and (2) of the theorem now follow from Theorems 3 and 2, respectively. $\square$

7 Typical behaviour of k -learnability

We now explore when one should expect multisets of randomly generated pure states to be k -learnable or not. Roughly speaking, there is a trade-off between the ambient dimension d and the number of states n : the more states are packed into smaller dimensions, the harder it is to k -learn those states.

Proposition 33. *Let $n \geq 2$, $k \in \{1, 2, \dots, n-1\}$, $0 < \delta < 1$, and let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be independent Haar-random pure states. If*

$$d < \frac{\pi}{4} \left(\frac{k-1}{n-1} + \sqrt{\frac{\ln(1/\delta)}{n-1}} \right)^{-2} \quad (111)$$

then these states are not k -learnable with probability at least $1 - \delta$. In particular, there is a constant $C_\delta > 0$ such that if $d \leq C_\delta \min\{n, (n/k)^2\}$ then these states are not k -learnable with probability at least $1 - \delta$.

Proof. Let G be the Gram matrix of the vectors $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle$. Then

$$\sum_{i,j} |\langle \psi_i | \psi_j \rangle| = n + \sum_{i \neq j} |\langle \psi_i | \psi_j \rangle|. \quad (112)$$

Let

$$U := \frac{1}{n(n-1)} \sum_{i \neq j} |\langle \psi_i | \psi_j \rangle|, \quad (113)$$

so that $\sum_{i,j} |\langle \psi_i | \psi_j \rangle| > nk$ is equivalent to $U > (k-1)/(n-1)$.

Recall that Hoeffding's inequality (see Inequality (5.7) in [Hoe63], for example) says that for every $t > 0$ we have

$$\Pr(U \leq \mu - t) \leq \exp(-2 \lfloor n/2 \rfloor t^2), \quad (114)$$

where $\mu = \mathbb{E}[U]$ (in the terminology and notation of [Hoe63], U is a “ U -statistic” associated with the function $g(|\psi_i\rangle, |\psi_j\rangle) = |\langle \psi_i | \psi_j \rangle|$).

Fix $i \neq j$. By unitary invariance of the Haar measure, all off-diagonal entries of G have the same expectation, so

$$\mu = \mathbb{E}[U] = \mathbb{E}[|\langle \psi_i | \psi_j \rangle|] = \frac{\sqrt{\pi}\Gamma(d)}{2\Gamma(d+1/2)} \geq \frac{\sqrt{\pi}}{2\sqrt{d}}, \quad (115)$$

where this value comes from [ZS00] and the inequality follows from Gautschi's inequality. We note that, by our hypothesis on d , we have

$$\mu \geq \frac{\sqrt{\pi}}{2\sqrt{d}} > \frac{k-1}{n-1} + \sqrt{\frac{\ln(1/\delta)}{n-1}} \geq \frac{k-1}{n-1}, \quad (116)$$

so we can choose

$$t = \mu - \frac{k-1}{n-1} > 0 \quad (117)$$

in Hoeffding's inequality.

With these values of μ and t , Inequality (114) tells us that

$$\begin{aligned} \Pr\left(U \leq \frac{k-1}{n-1}\right) &\leq \exp\left(-2 \left\lfloor \frac{n}{2} \right\rfloor \left(\mu - \frac{k-1}{n-1}\right)^2\right), \\ &\leq \exp\left(-(n-1) \left(\frac{\sqrt{\pi}}{2\sqrt{d}} - \frac{k-1}{n-1}\right)^2\right) \\ &\leq \delta, \end{aligned} \quad (118)$$

with the final inequality following from our hypothesis on d . It follows that

$$\Pr\left(\sum_{i,j} |\langle \psi_i | \psi_j \rangle| > nk\right) = \Pr\left(U > \frac{k-1}{n-1}\right) \geq 1 - \delta \quad (119)$$

which (thanks to Theorem 3) completes the proof. $\square$

The previous proposition showed that if d is small (compared to n) then we do not expect randomly generated states to be k -learnable. The following proposition shows that, by contrast, if d is large then we *do* expect randomly generated states to be k -learnable.

Proposition 34. *Let $n \geq 2$, $k \in \{2, \dots, n\}$, $0 < \delta < 1$, and let $S = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be independent Haar-random pure states. If*

$$d \geq \frac{(n-1)(n-k+1)}{\delta(k-1)} \quad (120)$$

then these states are k -learnable with probability at least $1 - \delta$. In particular, there is a constant $C_\delta > 0$ such that if $d \geq C_\delta(n(n-k+1)/k)$ then these states are k -learnable with probability at least $1 - \delta$.

Proof. Let G be the Gram matrix of the vectors $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle$. Then

$$\|G\|_F^2 = n + \sum_{i \neq j} |\langle \psi_i | \psi_j \rangle|^2. \quad (121)$$

Let

$$Y := \sum_{i \neq j} |\langle \psi_i | \psi_j \rangle|^2. \quad (122)$$

By Theorem 2 it suffices to prove that $Y \leq n(k-1)/(n-k+1)$ with probability at least $1 - \delta$.

Fix $i \neq j$. Since $|\psi_i\rangle$ and $|\psi_j\rangle$ are independent and Haar-random, we can write

$$\mathbb{E} \left[|\langle \psi_i | \psi_j \rangle|^2 \right] = \int \int |\langle \mathbf{x} | \mathbf{y} \rangle|^2 d\mu(\mathbf{y}) d\mu(\mathbf{x}), \quad (123)$$

where μ denotes Haar measure on the unit sphere in $\mathbb{C}^d$. Fix a unit vector $\mathbf{x} \in \mathbb{C}^d$. There exists a unitary matrix U such that $U\mathbf{x} = \mathbf{e}_1$ (the first standard basis vector). Using unitary invariance of the Haar measure,

$$\int |\langle \mathbf{x} | \mathbf{y} \rangle|^2 d\mu(\mathbf{y}) = \int |\langle U\mathbf{x} | U\mathbf{y} \rangle|^2 d\mu(\mathbf{y}) = \int |\langle \mathbf{e}_1 | \mathbf{y} \rangle|^2 d\mu(\mathbf{y}). \quad (124)$$

We thus see that the inner integral is independent of $\mathbf{x}$, so

$$\mathbb{E} \left[|\langle \psi_i | \psi_j \rangle|^2 \right] = \int |\langle \mathbf{e}_1 | \mathbf{y} \rangle|^2 d\mu(\mathbf{y}). \quad (125)$$

Write $\mathbf{y} = (\alpha_1, \alpha_2, \dots, \alpha_d)$. Then $\sum_{\ell=1}^d |\alpha_\ell|^2 = 1$, so (by unitary invariance of the Haar measure) all coordinates have the same expectation, which gives

$$\mathbb{E} \left[|\langle \psi_i | \psi_j \rangle|^2 \right] = \mathbb{E} \left[|\alpha_1|^2 \right] = \frac{1}{d}. \quad (126)$$

It follows that $\mathbb{E}[Y] = n(n-1)/d$. By Markov's inequality,

$$\Pr \left(Y > \frac{n(k-1)}{n-k+1} \right) \leq \frac{\mathbb{E}[Y]}{n(k-1)/(n-k+1)} = \frac{(n-1)(n-k+1)}{d(k-1)}. \quad (127)$$

The stated bound on d ensures that this quantity is at most δ , yielding the result. $\square$

Acknowledgements

J.S. thanks Richard Cleve and Ankith Mohan for helpful discussions. The authors used Fable-5 and GPT-5.6 Sol to prove Lemmas 35 and 36, which are central ingredients for the main result. Large language models were also used for proof ideas and editorial polishing throughout the paper. The authors take full responsibility for the correctness, exposition, and attribution in the final manuscript. N.J. acknowledges support from NSERC Discovery Grant number RGPIN-2022-04098. B.L. acknowledges support from NSERC Discovery Grant number RGPIN-2026-05413. J.S. acknowledges support from the National Science Foundation Award Number 2542721.

References

- [Aar07] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007.
- [Aar18] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 325–338, New York, NY, USA, 2018. Association for Computing Machinery.
- [ADIS25] Jonathan Allcock, Joao F Doriguello, Gábor Ivanyos, and Miklos Santha. Reconquering Bell sampling on qudits: stabilizer learning and testing, quantum pseudorandomness bounds, and more. *arXiv preprint arXiv:2510.06848*, 2025.
- [AG08] Scott Aaronson and Daniel Gottesman. Identifying stabilizer states. PIRSA talk, url: <http://pirsa.org/08080052/>, 2008.
- [BCPT05] Erik G. Boman, Doron Chen, Ojas Parekh, and Sivan Toledo. On factor width and symmetric H-matrices. *Linear Algebra and its Applications*, 405:239–248, 2005.
- [BDSS22] Grigoriy Blekherman, Santanu S. Dey, Kevin Shu, and Shengding Sun. Hyperbolic relaxation of k -locally positive semidefinite matrices. *SIAM Journal on Optimization*, 32(2):470–490, 2022.
- [BF03] John J. Benedetto and Matthew Fickus. Finite normalized tight frames. *Advances in Computational Mathematics*, 18(2-4):357–385, 2003.
- [BJOP14] Somshubhro Bandyopadhyay, Rahul Jain, Jonathan Oppenheim, and Christopher Perry. Conclusive exclusion of quantum states. *Physical Review A*, 89(2):022336, 2014.
- [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*, STOC 2019, pages 503–514, New York, NY, USA, 2019. Association for Computing Machinery.
- [CCLW26] Rebecca Chang, Matthias C. Caro, Martin Larocca, and Maxwell West. Sample-optimal learning of stabilizer states. *arXiv preprint arXiv:2609.10974*, 2026.
- [CFS02] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Conditions for compatibility of quantum-state assignments. *Physical Review A*, 66(6):062111, 2002.
- [Går59] Lars Gårding. An inequality for hyperbolic polynomials. *Journal of Mathematics and Mechanics*, 8(6):957–965, 1959.

- [GMC17] Héctor J García, Igor L Markov, and Andrew W Cross. On the geometry of stabilizer states. *arXiv preprint arXiv:1711.07848*, 2017.
- [Gro06] David Gross. Hudson’s theorem for finite-dimensional quantum systems. *Journal of Mathematical Physics*, 47(12):122107, 2006.
- [HB20] Vojtěch Havlíček and Jonathan Barrett. Simple communication complexity separation from quantum state antidistinguishability. *Physical Review Research*, 2(1):013326, 2020.
- [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.
- [HK18] Teiko Heinosaari and Oskari Kerppo. Antidistinguishability of pure quantum states. *Journal of Physics A: Mathematical and Theoretical*, 51(36):365303, 2018.
- [HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.
- [Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [JLRS25] Nathaniel Johnston, Benjamin Lovitz, Vincent Russo, and Jamie Sikora. The complexity of perfect quantum state classification. *arXiv preprint arXiv:2510.20789*, 2025.
- [JMPP22] Nathaniel Johnston, Shirin Moein, Rajesh Pereira, and Sarah Plosker. Absolutely k -incoherent quantum states and spectral inequalities for factor width of a matrix. *Physical Review A*, 106(5):052417, 2022.
- [JRS25] Nathaniel Johnston, Vincent Russo, and Jamie Sikora. Tight bounds for antidistinguishability and circulant sets of pure quantum states. *Quantum*, 9:1622, 2025.
- [KG15] Richard Kueng and David Gross. Qubit stabilizer states are complex projective 3-designs. *arXiv preprint arXiv:1510.02767*, 2015.
- [Lei14] Matthew Saul Leifer. Is the quantum state real? An extended review of ψ -ontology theorems. *Quanta*, 3(1):67–155, 2014.
- [LR18] Nana Liu and Patrick Rebentrost. Quantum machine learning for quantum anomaly detection. *Physical Review A*, 97:042315, 2018.
- [LSMT24] Santiago Llorens, Gael Sentís, and Ramon Muñoz-Tapia. Quantum multi-anomaly detection. *Quantum*, 8:1452, 2024.
- [MNW24] Hemant K. Mishra, Michael Nussbaum, and Mark M. Wilde. On the optimal error exponents for classical and quantum antidistinguishability. *Letters in Mathematical Physics*, 114(3):76, 2024.
- [Mon17] Ashley Montanaro. Learning stabilizer states by Bell sampling. *arXiv preprint arXiv:1707.04012*, 2017.
- [MSU23] Ankith Mohan, Jamie Sikora, and Sarvagya Upadhyay. A generalized framework for quantum state discrimination, hybrid algorithms, and the quantum change point problem. *arXiv preprint arXiv:2312.04023v1*, 2023.

- [OHSG24] Valentin Obst, Arne Heimendahl, Tanmay Singal, and David Gross. Wigner’s theorem for stabilizer states and quantum designs. *Journal of Mathematical Physics*, 65:112202, 2024.
- [OW21] Ryan O’Donnell and John Wright. Quantum spectrum testing. *Communications in Mathematical Physics*, 387(1):1–75, 2021.
- [PBR12] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, 2012.
- [RBC⁺18] M. Ringbauer, T. R. Bromley, M. Cianciaruso, L. Lami, W. Y. S. Lau, G. Adesso, A. G. White, A. Fedrizzi, and M. Piani. Certification and quantification of multilevel quantum coherence. *Physical Review X*, 8(4):041007, 2018.
- [RS23] Vincent Russo and Jamie Sikora. Inner products of pure states and their antidistinguishability. *Physical Review A*, 107(3):L030202, 2023.
- [SBC⁺16] Gael Sentís, Emilio Bagan, John Calsamiglia, Giulio Chiribella, and Ramon Muñoz-Tapia. Quantum change point. *Physical Review Letters*, 117(15):150502, 2016.
- [SCMT17] Gael Sentís, John Calsamiglia, and Ramon Muñoz-Tapia. Exact identification of a quantum change point. *Physical Review Letters*, 119(14):140506, 2017.
- [Wel74] Lloyd R. Welch. Lower bounds on the maximum cross correlation of signals. *IEEE Transactions on Information Theory*, 20(3):397–399, 1974.
- [ŻS00] Karol Życzkowski and Hans-Jürgen Sommers. Truncations of random unitary matrices. *Journal of Physics A: Mathematical and General*, 33(10):2045–2057, 2000.

A Deferred proofs

A.1 Proof of Proposition 4

Proof. Let G be the Gram matrix of S . Suppose first that $R_i \leq 2$ for all i . Since $G_{i,i} = 1$ for every i ,

$$\sum_{j \neq i} |G_{i,j}| = R_i - 1 \leq 1. \quad (128)$$

Thus G is diagonally dominant. We can write

$$G = \sum_{i=1}^n \left(1 - \sum_{j \neq i} |G_{i,j}| \right) \mathbf{e}_i \mathbf{e}_i^* + \sum_{1 \leq i < j \leq n} B^{i,j}, \quad (129)$$

where $B^{i,j}$ is supported only on rows and columns i, j and has principal 2×2 block

$$\begin{pmatrix} |G_{i,j}| & G_{i,j} \\ \overline{G_{i,j}} & |G_{i,j}| \end{pmatrix}. \quad (130)$$

Note that each $B^{i,j}$ is positive semidefinite. Hence G is a sum of positive semidefinite matrices supported on at most two coordinates, so it is 2-incoherent. By Proposition 10, S is 2-learnable.

Now suppose that S is 2-learnable. By Theorem 3,

$$\sum_{i,j=1}^n |G_{i,j}| \leq 2n, \quad (131)$$

or equivalently, $\sum_{i=1}^n R_i \leq 2n$. □

A.2 Proof of Theorem 2

Define $r := n - k + 1$. For $\mathbf{d} \in \mathbb{R}^n$, let $e_j(\mathbf{d})$ denote the j -th elementary symmetric polynomial of the coordinates of $\mathbf{d}$:

$$e_j(\mathbf{d}) := \sum_{1 \leq i_1 < \dots < i_j \leq n} \left(\prod_{\ell=1}^j d_{i_\ell} \right), \quad (132)$$

with the convention that $e_0(\mathbf{d}) = 1$. Define the cone

$$\Gamma_k^{(n)} := \{\mathbf{d} \in \mathbb{R}^n : e_1(\mathbf{d}) \geq 0, \dots, e_k(\mathbf{d}) \geq 0\}, \quad (133)$$

and the compact convex set

$$\Delta_r^{(n)} := \left\{ \mathbf{p} \in \mathbb{R}_{\geq 0}^n : \sum_{j=1}^n p_j = 1, \|\mathbf{p}\|_2 \leq \frac{1}{\sqrt{r}} \right\}. \quad (134)$$

It is known [Går59] that $\Gamma_k^{(n)}$ is a closed convex cone and that $\mathbb{R}_{\geq 0}^n \subseteq \Gamma_k^{(n)}$. We use $(\Gamma_k^{(n)})^\circ$ to denote the dual cone of $\Gamma_k^{(n)}$:

$$(\Gamma_k^{(n)})^\circ := \{\mathbf{y} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{y} \geq 0 \text{ for all } \mathbf{x} \in \Gamma_k^{(n)}\}. \quad (135)$$

Before proving the result, we need a pair of helper lemmas.

Lemma 35. *Let $n \geq 4$ be an integer and let $k \in \{3, 4, \dots, n-1\}$. Define $r := n - k + 1$ and suppose $\mathbf{p} \in \mathbb{R}^n$ satisfies $\|\mathbf{p}\|_2^2 = 1/r$, $\sum_{i=1}^n p_i = 1$, and $p_i > 0$ for all $i \in \{1, 2, \dots, n\}$. Then $\mathbf{1} - r\mathbf{p} \notin \Gamma_k^{(n)}$.*

Proof. Define $\mathbf{x} := \mathbf{1} - r\mathbf{p}$ and suppose (for the purpose of establishing a contradiction) that $\mathbf{x} \in \Gamma_k^{(n)}$. Notice that $e_1(\mathbf{x}) = \sum_{i=1}^n x_i = n - r = k - 1$ and

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n (1 - rp_i)^2 = \sum_{i=1}^n (1 - 2rp_i + r^2 p_i^2) = n - 2r + r = n - r = k - 1. \quad (136)$$

Our first goal is to prove that

$$ke_k(\mathbf{x}) = -r \sum_{i=1}^n p_i x_i^2 e_{k-3}(\mathbf{x}_{\hat{i}}), \quad (137)$$

where $\mathbf{x}_{\hat{i}} \in \mathbb{R}^{n-1}$ is the vector $\mathbf{x}$ with the i -th entry removed. To this end, recall the standard identities

$$ke_k(\mathbf{x}) = \sum_{i=1}^n x_i e_{k-1}(\mathbf{x}_{\hat{i}}) \quad \text{and} \quad e_{k-1}(\mathbf{x}_{\hat{i}}) = e_{k-1}(\mathbf{x}) - x_i e_{k-2}(\mathbf{x}_{\hat{i}}) \quad (138)$$

for all $i \in \{1, 2, \dots, n\}$. From here, we have

$$ke_k(\mathbf{x}) = \sum_{i=1}^n x_i e_{k-1}(\mathbf{x}_{\hat{i}}) \quad (139)$$

$$= \sum_{i=1}^n x_i (e_{k-1}(\mathbf{x}) - x_i e_{k-2}(\mathbf{x}_{\hat{i}})) \quad (140)$$

$$= (k-1)e_{k-1}(\mathbf{x}) - \sum_{i=1}^n x_i^2 e_{k-2}(\mathbf{x}_{\hat{i}}) \quad (141)$$

$$= -\sum_{i=1}^n (x_i^2 - x_i) e_{k-2}(\mathbf{x}_{\hat{i}}). \quad (142)$$

Note that $x_i^2 - x_i = -rp_i x_i$, so the above string of equations becomes

$$ke_k(\mathbf{x}) = r \sum_{i=1}^n p_i x_i e_{k-2}(\mathbf{x}_{\hat{i}}). \quad (143)$$

Using the recursion $e_{k-2}(\mathbf{x}_{\hat{i}}) = e_{k-2}(\mathbf{x}) - x_i e_{k-3}(\mathbf{x}_{\hat{i}})$, we obtain

$$ke_k(\mathbf{x}) = r \sum_{i=1}^n p_i x_i (e_{k-2}(\mathbf{x}) - x_i e_{k-3}(\mathbf{x}_{\hat{i}})). \quad (144)$$

Now observe that

$$\sum_{i=1}^n p_i x_i = \sum_{i=1}^n p_i (1 - rp_i) = \sum_{i=1}^n p_i - r \sum_{i=1}^n p_i^2 = 1 - \frac{r}{r} = 0. \quad (145)$$

Substituting into Equation (144) (finally) gives

$$ke_k(\mathbf{x}) = -r \sum_{i=1}^n p_i x_i^2 e_{k-3}(\mathbf{x}_{\hat{i}}), \quad (146)$$

as desired.

By [JMPP22, Lemma 2], deleting one coordinate from a vector in $\Gamma_k^{(n)}$ produces a vector in $\Gamma_{k-1}^{(n-1)}$. It follows that every term on the right-hand side of (146) (ignoring the leading negative sign) is nonnegative. It follows that $e_k(\mathbf{x}) \leq 0$. Since $\mathbf{x} \in \Gamma_k^{(n)}$, we must actually have $e_k(\mathbf{x}) = 0$, from which Equation (146) implies

$$e_{k-3}(\mathbf{x}_{\hat{i}}) = 0 \quad \text{whenever} \quad x_i \neq 0. \quad (147)$$

Now let $\mathbf{y} \in \mathbb{R}^m$ be the vector of nonzero entries of $\mathbf{x}$. Every zero entry of $\mathbf{x}$ must be in the same location as an entry of $\mathbf{p}$ that is equal to $1/r$. Since $\mathbf{p}$ has full support, $\mathbf{x}$ has at most $r - 1$ zero entries, so $m \geq n - r + 1 = k$. Equation (147) therefore says that $e_{k-3}(\mathbf{y}_{\hat{i}}) = 0$ for all $i \in \{1, 2, \dots, m\}$.

This is impossible. Indeed, it is not possible that $y_i \neq 0$ for all $i \in \{1, 2, \dots, m\}$ and there exists $0 \leq \ell \leq m - 1$ for which $e_\ell(\mathbf{y}_{\hat{i}}) = 0$ for all $i \in \{1, 2, \dots, m\}$. The $\ell = 0$ case is not possible simply because $e_0(\mathbf{y}_{\hat{i}}) = 1$. For $\ell \geq 1$, recall again the identity

$$e_\ell(\mathbf{y}) = e_\ell(\mathbf{y}_{\hat{i}}) + y_i e_{\ell-1}(\mathbf{y}_{\hat{i}}) \quad (148)$$

for all $i \in \{1, 2, \dots, m\}$. Since $e_\ell(\mathbf{y}_{\hat{i}}) = 0$ by assumption, this becomes $e_\ell(\mathbf{y}) = y_i e_{\ell-1}(\mathbf{y}_{\hat{i}})$. Summing over all i gives $me_\ell(\mathbf{y}) = \ell e_\ell(\mathbf{y})$, so $e_\ell(\mathbf{y}) = 0$ and then $e_{\ell-1}(\mathbf{y}_{\hat{i}}) = 0$ for every i . Descending in ℓ eventually gives $e_0(\mathbf{y}_{\hat{i}}) = 0$, contradicting $e_0(\mathbf{y}_{\hat{i}}) = 1$ and proving the lemma. $\square$

Lemma 36. Suppose $n \geq 2$ is an integer and $k \in \{1, 2, \dots, n - 1\}$, and let $r := n - k + 1$. Then $\Delta_r^{(n)} \subseteq (\Gamma_k^{(n)})^\circ$.

Proof. We need to show that if $\mathbf{d} \in \Gamma_k^{(n)}$ and $\mathbf{p} \in \Delta_r^{(n)}$ then $\mathbf{d} \cdot \mathbf{p} \geq 0$. We use induction on k . For $k = 1$, the Cauchy–Schwarz inequality forces every $\mathbf{p} \in \Delta_n^{(n)}$ to equal $\mathbf{1}/n$, so $\mathbf{d} \cdot \mathbf{p} = e_1(\mathbf{d})/n \geq 0$. For $k = 2$, the result is exactly the description of $(\Gamma_2^{(n)})^\circ$ from [JMPP22, Theorem 6].

Now let $k \geq 3$, fix $\mathbf{d} \in \Gamma_k^{(n)}$, and let $\mathbf{p}^*$ minimize $\mathbf{d} \cdot \mathbf{p}$ over $\Delta_r^{(n)}$. Put $I := \text{supp}(\mathbf{p}^*)$ and $m := |I|$. The Cauchy–Schwarz inequality gives $m \geq r$. We now split into two cases depending on the size of m :

Case 1: $m < n$. Let $\mathbf{d}_I$ and $\mathbf{p}_I^*$ denote the restrictions of $\mathbf{d}$ and $\mathbf{p}^*$, respectively, to the entries indexed by I . Define $k' := m - r + 1 = k - (n - m)$. Iterating [JMPP22, Lemma 2] gives $\mathbf{d}_I \in \Gamma_{k'}^{(m)}$, while $\mathbf{p}_I^* \in \Delta_r^{(m)}$ and $r = m - k' + 1 \geq 2$, so $m > k'$. Since $1 \leq k' < k$, the inductive hypothesis yields

$$\mathbf{d} \cdot \mathbf{p}^* = \mathbf{d}_I \cdot \mathbf{p}_I^* \geq 0. \quad (149)$$

Case 2: $m = n$. Then $p_i^* > 0$ for all i , so the nonnegativity constraints are inactive. Since $\mathbf{1}/n$ is strictly feasible, the Karush–Kuhn–Tucker conditions give numbers $\tau \in \mathbb{R}$ and $s \geq 0$ such that

$$\mathbf{d} = \tau \mathbf{1} - s \mathbf{p}^*. \quad (150)$$

If $s = 0$ then $\mathbf{d} = \tau \mathbf{1}$, so $n\tau = e_1(\mathbf{d}) \geq 0$ implies $\mathbf{d} \cdot \mathbf{p}^* = \tau \geq 0$.

On the other hand, if $s > 0$ then $n\tau - s = e_1(\mathbf{d}) \geq 0$ implies $\tau > 0$. Assume (for the sake of establishing a contradiction) that $\mathbf{d} \cdot \mathbf{p}^* < 0$. Then the fact that $\|\mathbf{p}^*\|_2^2 = 1/r$ implies

$$0 > \mathbf{d} \cdot \mathbf{p}^* = \tau - \frac{s}{r}, \quad (151)$$

so $\beta := s/\tau > r$. Since $\mathbf{1} - \beta \mathbf{p}^* = \mathbf{d}/\tau \in \Gamma_k^{(n)}$ and $\mathbf{1} \in \Gamma_k^{(n)}$, convexity of $\Gamma_k^{(n)}$ gives

$$\mathbf{1} - r \mathbf{p}^* = \left(1 - \frac{r}{\beta}\right) \mathbf{1} + \frac{r}{\beta} (\mathbf{1} - \beta \mathbf{p}^*) \in \Gamma_k^{(n)}, \quad (152)$$

contradicting Lemma 35. It follows that $\mathbf{d} \cdot \mathbf{p}^* \geq 0$, which completes the induction and the proof. $\square$

Proof of Theorem 2. The theorem is trivial when $n \leq 2$ or $k \in \{1, n\}$, so we assume throughout the proof that $n \geq 3$ and $k \in \{2, 3, \dots, n-1\}$. Let G be a Gram matrix with $\|G\|_F \leq n/\sqrt{r}$ and define the matrix $X := G/n$. Then X is a density matrix (positive semidefinite with trace 1) and

$$\|X\|_F^2 = \frac{\|G\|_F^2}{n^2} \leq \frac{1}{r}. \quad (153)$$

It follows that $\mathbf{p} := \lambda(X) \in \Delta_r^{(n)}$. By Lemma 36, this implies $\mathbf{p} \in \left(\Gamma_k^{(n)}\right)^\circ$. It follows from [JMPP22, Theorem 3] that X is k -incoherent, so the same is true of G . By Proposition 10, we conclude that every multiset of states with Gram matrix G is k -learnable, which completes the proof. $\square$

A.3 Proof of Theorem 3

Proof. We prove the contrapositive: if S is k -learnable then

$$\sum_{i,j=1}^n |\langle \psi_i | \psi_j \rangle| \leq nk. \quad (154)$$

To this end, let G be the Gram matrix of S and define

$$Y = k\mathbf{1} - E, \quad (155)$$

where, for all $1 \leq i, j \leq n$ (even if $i = j$), the (i, j) -entry of E is the complex number with modulus 1 and phase equal to that of $\langle \psi_i | \psi_j \rangle$ (if $\langle \psi_i | \psi_j \rangle = 0$ then set $E_{i,j} = 1$). Each $k \times k$ principal submatrix of Y is diagonally dominant and thus positive semidefinite. It follows that, in the terminology of [BDSS22, JMPP22], Y is “ k -locally positive semidefinite” and thus in the dual cone of the set of k -incoherent matrices. Since G is k -incoherent by Proposition 10, this tells us that $\text{Tr}(YG) \geq 0$. Direct calculation then shows that

$$0 \leq \text{Tr}(YG) = \text{Tr}((k\mathbb{1} - E)G) = n(k-1) - \sum_{\substack{i,j=1 \\ i \neq j}}^n |\langle \psi_i | \psi_j \rangle|, \quad (156)$$

which is equivalent to Inequality (154). This completes the proof. $\square$

A.4 Proof of Proposition 24

Proof. Fix $c \geq 1$ and define

$$h_{d,m} := \left\| G_{d,m}^{(c)} \right\|_F^2 = \sum_{\psi, \phi \in \text{Stab}_{d,m}} |\langle \psi | \phi \rangle|^{2c}. \quad (157)$$

Let $\mathcal{F}_c(d, m)$ be the c -th *frame potential* of the uniform distribution over $\text{Stab}_{d,m}$, defined by

$$h_{d,m} = s_{d,m}^2 \mathcal{F}_c(d, m). \quad (158)$$

By [KG15, Theorem 2], we have

$$\frac{\mathcal{F}_c(d, m+1)}{\mathcal{F}_c(d, m)} = \frac{d^{m-(c-2)} + 1}{d(d^{m+1} + 1)}. \quad (159)$$

Note that

$$\frac{s_{d,m+1}}{s_{d,m}} = d(d^{m+1} + 1). \quad (160)$$

Using (158), (159) and (160) we compute

$$\begin{aligned} \frac{h_{d,m+1}}{h_{d,m}} &= \left(\frac{s_{d,m+1}}{s_{d,m}} \right)^2 \frac{\mathcal{F}_c(d, m+1)}{\mathcal{F}_c(d, m)} \\ &= \left(d(d^{m+1} + 1) \right)^2 \frac{d^{m-(c-2)} + 1}{d(d^{m+1} + 1)} = d(d^{m+1} + 1) \left(d^{m-(c-2)} + 1 \right). \end{aligned} \quad (161)$$

Define $t_{d,m} := h_{d,m} / s_{d,m}$ so that

$$t_{d,m+1} = \frac{h_{d,m+1}}{s_{d,m+1}} = \frac{h_{d,m} d(d^{m+1} + 1) (d^{m-(c-2)} + 1)}{s_{d,m} d(d^{m+1} + 1)} = t_{d,m} (d^{m-(c-2)} + 1). \quad (162)$$

For the base case $m = 1$, [KG15, Theorem 2] gives

$$\mathcal{F}_c(d, 1) = \frac{d^{2-c} + 1}{d(d+1)}. \quad (163)$$

Since $s_{d,1} = d(d+1)$, we obtain

$$h_{d,1} = s_{d,1}^2 \mathcal{F}_c(d, 1) = (d(d+1))^2 \frac{d^{2-c} + 1}{d(d+1)} = d(d+1) (d^{2-c} + 1), \quad (164)$$

hence

$$t_{d,1} = \frac{h_{d,1}}{s_{d,1}} = d^{2-c} + 1. \quad (165)$$

For $m \geq 2$, iterating (162) yields

$$t_{d,m} = (d^{2-c} + 1) \prod_{j=1}^{m-1} (1 + d^{j-(c-2)}) = \prod_{j=0}^{m-1} (1 + d^{j-(c-2)}), \quad (166)$$

which gives

$$h_{d,m} = s_{d,m} t_{d,m} = s_{d,m} \prod_{j=0}^{m-1} (1 + d^{j-(c-2)}), \quad (167)$$

as claimed. $\square$

A.5 Proof of Corollary 27

Proof. By Theorem 26, $\text{Stab}_{d,m}$ is $(c, 2)$ -learnable if and only if

$$\prod_{j=0}^{m-1} (1 + d^{j+2-c/2}) \leq 2. \quad (168)$$

First suppose $m = 1$. For $c = 4$ the left-hand side is $1 + d^{2-4/2} = 2$, whereas for $c = 3$ it is $1 + d^{1/2} > 2$, so $c_2(\text{Stab}_{d,1}) = 4$ for every prime d .

Now suppose $m \geq 2$. For $c = 2m + 2$,

$$\prod_{j=0}^{m-1} (1 + d^{j+2-c/2}) = \prod_{r=1}^m (1 + d^{1-r}). \quad (169)$$

The factor corresponding to $r = 1$ equals 2, while all remaining factors are strictly greater than 1. It follows that

$$\prod_{r=1}^m (1 + d^{1-r}) > 2, \quad (170)$$

and therefore $c_2(\text{Stab}_{d,m}) \geq 2m + 3$ for all $m \geq 2$.

We now split into cases to consider the different possible local dimensions d .

Case 1: $d \geq 5$. Set $c = 2m + 3$. Then

$$\prod_{j=0}^{m-1} (1 + d^{j+2-c/2}) = \prod_{r=1}^m (1 + d^{-r+1/2}). \quad (171)$$

Using the inequality $\ln(1 + x) \leq x$, we see that

$$\ln \left(\prod_{r=1}^m (1 + d^{-r+1/2}) \right) \leq \sum_{r=1}^m d^{-r+1/2} = \frac{\sqrt{d}}{d-1} \leq \frac{\sqrt{5}}{4} < \ln(2). \quad (172)$$

It follows that

$$\prod_{r=1}^m (1 + d^{-r+1/2}) < 2, \quad (173)$$

so $2m + 3$ copies suffice. Since we already showed that $c_2(\text{Stab}_{d,m}) \geq 2m + 3$, we conclude that $c_2(\text{Stab}_{d,m}) = 2m + 3$ whenever $d \geq 5$ and $m \geq 2$.

Case 2: $d = 3$. For $c = 2m + 3$, we have

$$\prod_{j=0}^{m-1} \left(1 + 3^{j+2-c/2}\right) = \prod_{r=1}^m \left(1 + 3^{-r+1/2}\right). \quad (174)$$

When $m = 2$, this product becomes

$$\left(1 + \frac{1}{\sqrt{3}}\right) \left(1 + \frac{1}{3\sqrt{3}}\right) < 2, \quad (175)$$

so 7 copies suffice. Since we already showed that $c_2(\text{Stab}_{3,2}) \geq 2m + 3 = 7$, we conclude that $c_2(\text{Stab}_{3,2}) = 7$.

On the other hand, when $m \geq 3$, this product can be bounded by

$$\prod_{r=1}^m \left(1 + 3^{-r+1/2}\right) \geq \left(1 + \frac{1}{\sqrt{3}}\right) \left(1 + \frac{1}{3\sqrt{3}}\right) \left(1 + \frac{1}{9\sqrt{3}}\right) > 2, \quad (176)$$

so $2m + 3$ copies do not suffice. On the other hand, for $c = 2m + 4$,

$$\prod_{j=0}^{m-1} \left(1 + 3^{j+2-c/2}\right) = \prod_{r=1}^m (1 + 3^{-r}), \quad (177)$$

and

$$\ln \left(\prod_{r=1}^m (1 + 3^{-r}) \right) \leq \sum_{r=1}^{\infty} 3^{-r} = \frac{1}{2} < \ln(2), \quad (178)$$

so $2m + 4$ copies suffice. It follows that $c_2(\text{Stab}_{3,m}) = 2m + 4$ whenever $m \geq 3$.

Case 3: $d = 2$. For $m = 2$, direct substitution into Theorem 26 gives $c_2(\text{Stab}_{2,2}) = 8$. Now suppose $m \geq 3$. If $c = 2m + 4$, then

$$\prod_{j=0}^{m-1} \left(1 + 2^{j+2-c/2}\right) = \prod_{r=1}^m (1 + 2^{-r}) \geq \left(1 + \frac{1}{2}\right) \left(1 + \frac{1}{4}\right) \left(1 + \frac{1}{8}\right) > 2. \quad (179)$$

It follows that $2m + 4$ copies do not suffice.

For $c = 2m + 5$,

$$\prod_{j=0}^{m-1} \left(1 + 2^{j+2-c/2}\right) = \prod_{r=1}^m \left(1 + 2^{-r-1/2}\right) < \prod_{r=1}^{\infty} \left(1 + 2^{-r-1/2}\right). \quad (180)$$

Again using $\ln(1 + x) \leq x$, we obtain

$$\prod_{r=1}^{\infty} \left(1 + 2^{-r-1/2}\right) = \left(1 + 2^{-3/2}\right) \prod_{r=2}^{\infty} \left(1 + 2^{-r-1/2}\right) < 2, \quad (181)$$

so $2m + 5$ copies suffice. It follows that $c_2(\text{Stab}_{2,m}) = 2m + 5$ whenever $m \geq 3$, which completes the proof. $\square$