

The complexity of perfect quantum state classification

Nathaniel Johnston*, Benjamin Lovitz†, Vincent Russo‡, Jamie Sikora§

October 24, 2025

Abstract

The problem of quantum state classification asks how accurately one can identify an unknown quantum state that is promised to be drawn from a known set of pure states. In this work, we introduce the notion of k -learnability, which captures the ability to identify the correct state using at most k guesses, with zero error. We show that deciding whether a given family of states is k -learnable can be solved via semidefinite programming. When there are n states, we present polynomial-time (in n) algorithms for determining k -learnability for two cases: when k is a fixed constant or the dimension of the states is a fixed constant. When both k and the dimension of the states are part of the input, we prove that there exist succinct certificates placing the problem in NP, and we establish NP-hardness by a reduction from the classical k -clique problem. Together, our findings delineate the boundary between efficiently solvable and intractable instances of quantum state classification in the perfect (zero-error) regime.

1 Introduction

Quantum state distinguishability is a cornerstone problem in quantum information theory [1–9] that addresses a fundamental question: *Given a quantum system prepared in one of several known states, can we identify the given state?* This problem has deep connections to quantum communication, cryptography, and computational complexity [10–13].

In the setting of quantum state exclusion [14], the goal is to determine which state the system is *not* in, rather than which state it *is*. Perfect quantum state exclusion (also called antidistinguishability [15–20]) of a state from a known set is possible when there exists a measurement that perfectly *rules out* a state that was sent. This exclusion perspective has proven fruitful for understanding nonlocal properties of quantum states [21–24] and has applications in quantum cryptographic protocols [25]. One noteworthy point (that we leverage later in this paper) is that the notion of antidistinguishability and the property of k -incoherence of a matrix [26] are intricately linked [20].

This work explores intermediate scenarios between perfect distinguishability and perfect exclusion. Rather than asking whether we can identify or rule out states with certainty, we ask: *How many guesses are needed to guarantee correct identification?* This perspective naturally leads to a new task which we call k -learnability: an ensemble of states is k -learnable if there exists a measurement

*Department of Mathematics & Computer Science, Mount Allison University.

njohnston@mta.ca, <https://njohnston.ca/>

†Department of Computer Science and Software Engineering, Concordia University.

benjamin.lovitz@concordia.ca, <https://www.benjaminlovitz.com/>

‡Unitary Foundation.

vincent@unitary.foundation, <https://vprusso.github.io/>

§Department of Computer Science, Virginia Tech.

sikora@vt.edu, <https://sites.google.com/site/jamiesikora/>

that, upon obtaining an outcome, allows us to narrow down the identity of the unknown state to at most k possibilities, with the guarantee that one of these k guesses is correct.

The k -learnability problem has natural connections to quantum machine learning [27, 28], quantum learning theory [29, 30], and tomography [31, 32]. In these fields, the problem of learning quantum states from measurement data has received considerable attention, with complexity measures often tied to the number of measurements or samples required. The k -learnability framework reinterprets learnability: it measures the number of distinct candidate states that must be considered, rather than the number of identical samples required.

1.1 Perfectly distinguishability and learning width

An indexed list of (pure) states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ is called *perfectly distinguishable* if there exists a POVM that, when applied to an unknown state drawn from the set, always perfectly guesses the index (that is, without error). It is well-known that a set of quantum states is perfectly distinguishable if and only if the states are pairwise orthogonal. However, if we relax this definition, one can, perhaps surprisingly, learn other aspects of this set without error, even when the states have non-trivial overlap.

The task of perfect state exclusion (also called antidistinguishability) asks when it is possible to perfectly *exclude* an index. A popular example where this can be achieved is the so-called trine states [33, 34]

$$|\psi_1\rangle = |0\rangle, \quad |\psi_2\rangle = -\frac{1}{2}(|0\rangle + \sqrt{3}|1\rangle), \quad |\psi_3\rangle = -\frac{1}{2}(|0\rangle - \sqrt{3}|1\rangle). \quad (1)$$

Optimal bounds for this task are known [14, 20] noting that one need not require pairwise orthogonality for a set of states to be antidistinguishable.

To illustrate k -learnability, consider the following generalization of the trine states to four states in three dimensions, called the *tetrahedral states*:

$$\begin{aligned} |\psi_1\rangle &= \frac{1}{\sqrt{3}}(|0\rangle + |1\rangle + |2\rangle), & |\psi_2\rangle &= \frac{1}{\sqrt{3}}(|0\rangle - |1\rangle - |2\rangle), \\ |\psi_3\rangle &= \frac{1}{\sqrt{3}}(-|0\rangle - |1\rangle + |2\rangle), & |\psi_4\rangle &= \frac{1}{\sqrt{3}}(-|0\rangle + |1\rangle - |2\rangle). \end{aligned} \quad (2)$$

These states are *2-learnable*: upon receiving one of them, one can always guess two states from which it was selected without error. This can be accomplished using the POVM $M_{i,j} = \frac{1}{2}|\phi_{i,j}\rangle\langle\phi_{i,j}|$, where

$$\begin{aligned} |\phi_{1,2}\rangle &= \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle), & |\phi_{1,3}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), & |\phi_{1,4}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |2\rangle), \\ |\phi_{2,3}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |2\rangle), & |\phi_{2,4}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), & |\phi_{3,4}\rangle &= \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle). \end{aligned} \quad (3)$$

We now formally define the notion of k -learnability.

Definition 1. Let k be a positive integer. An indexed list of (possibly sub-normalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ is k -learnable if there exists a POVM

$$\{M_S : S \subseteq \{1, \dots, n\}, |S| = k\}, \quad (4)$$

such that $\langle\psi_i|M_S|\psi_i\rangle = 0$ whenever $i \notin S$. The learning width of the list is the smallest k such that it is k -learnable.

Remark 2. We note that later, when we discuss the decision problem of determining the k -learnability of pure states, we need to define approximate k -learnability. For this, it is convenient to define the k -learnability of subnormalized pure states. Notice that as long as the states have positive norm, the task of determining if they are k -learnable is the same; it is only a matter of finding a POVM with the correct indices and orthogonality conditions.

Note that 1-learnability is equivalent to perfect distinguishability and $(n - 1)$ -learnability is equivalent to antidistinguishability. A set has learning width 1 if and only if the states are pairwise orthogonal. Every set has learning width at most n , trivially, and a set is antidistinguishable if and only if it has learning width at most $n - 1$. The trine states and the tetrahedral states both have learning width 2 since they are 2-learnable but not 1-learnable.

We can trivially construct indexed lists of pure states that are k -learnable for any $n \geq k$ in the following way. Consider the indexed list of states

$$\{\underbrace{|0\rangle, |0\rangle, \dots, |0\rangle}_{k \text{ copies}}, \underbrace{|1\rangle, |1\rangle, \dots, |1\rangle}_{k \text{ copies}}, |2\rangle, |2\rangle, \dots\}. \quad (5)$$

Measuring in the computational basis will reveal which k -subset the states were in. While this example is rather straightforward, we note that it also illustrates why we use the term *indexed list* of states instead of *set*, as we are guessing the index, not the state. Indeed, with the states in Equation (5), one can always guess the state, but not necessarily the index.

1.2 Our results

We establish both positive and negative results concerning the computational complexity of deciding k -learnability. Specifically, we show that when certain parameters are fixed, the problem admits efficient algorithms: the k -learnability of a set of n states can be decided in $\text{poly}(n)$ time when either the number of guesses k is fixed (Theorem 3) or when the dimension of the underlying Hilbert space is fixed (Theorem 4). In these regimes, the structure of the problem allows reductions to semidefinite programs of bounded size, which can be solved efficiently using standard approximation techniques. In contrast, when both k and the dimension are treated as part of the input, the problem becomes substantially harder: we prove that it is NP-complete by establishing membership in NP and constructing a reduction from the classical k -clique problem (Theorem 5, Theorem 6). Together, these results delineate the boundary between the tractable and intractable instances of quantum state classification. In what follows, we state these results informally and provide intuition for the underlying proofs.

The starting point for our proofs is the observation that an indexed list of states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ is k -learnable if and only if its associated Gram matrix $G_{ij} = \langle \psi_i | \psi_j \rangle$ possesses a particular structural property called k -incoherence [26]. Intuitively, this means that the Gram matrix can be expressed as a convex combination of outer products of vectors that are supported on at most k coordinates. In the following sections (see Section 2.1), we formalize this correspondence and show how it serves as the foundation for both our algorithmic and hardness results.

Theorem 3 (Informal, see Section 6). *Given an indexed list of (possibly sub-normalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^n$, there exists a polynomial-time algorithm (in n) to test if the list is k -learnable when k is a fixed constant.*

The proof proceeds by stating the problem as a semidefinite program in which there are polynomially-bounded variable size and also polynomially-bounded number of constraints. Then

using standard algorithms for efficiently approximating semidefinite programs, e.g., the Ellipsoid Method [35], this gives the result.

The next theorem gives another regime where we can efficiently test k -learnability.

Theorem 4 (Informal, see Corollary 29). *Given an indexed list of (possibly sub-normalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$, there exists a polynomial-time algorithm to test if the list is k -learnable when k is now part of the input, but the dimension d is fixed.*

Translating to k -incoherence, this result says that it can be determined in polynomial time whether the (constant) rank d Gram matrix is k -incoherent. This is clear when $d = 1$ by extremality of pure states. In general, one can formulate the question of k -incoherence as a semidefinite program in $\sim \binom{n}{k}$ matrix variables, which could be exponential in n , e.g., if $k \approx n/2$. The crux of the proof is the observation that for any constant rank, one can efficiently discard all but $\text{poly}(n)$ of these matrix variables.

When the number of allowed guesses k is treated as part of the input, and the Hilbert space dimension is also allowed to scale with the problem size, the situation changes dramatically. In this more general setting, the combinatorial structure of the k -learnability condition becomes significantly richer, and the semidefinite formulations that are efficient in the fixed-parameter cases no longer suffice to yield a polynomial-time algorithm. Intuitively, as both k and the dimension grow, the space of possible candidate decompositions of the Gram matrix expands exponentially, making it (seemingly) computationally infeasible to solve via convex optimization. In fact, we show that the problem is unlikely to be tractable by showing that it is NP-complete. This is established through two complementary results, one demonstrating that k -learnability lies in NP via succinct proofs, and another proving NP-hardness by reduction from the classical k -clique problem. Together, these theorems provide a tight characterization of the problem's computational difficulty in the general case.

Theorem 5 (Informal, see Section 5). *The decision variant of k -learnability is in NP.*

We introduce the decision variant of the problem in Section 3.

Notice that a proof of k -learnability could be to simply non-deterministically guess the POVM operators in the learning POVM. However, there are $\binom{n}{k}$ such operators, which is exponentially large when $k \approx n/2$ (as seen by Stirling's approximation, for example), thus a different method is required. To do this, we reformulate the problem in terms of k -incoherence, and use the geometry of that set along with Carathéodory's theorem. This shows the existence of succinct proofs.

Theorem 6 (Informal, see Theorem 21). *The decision variant of k -learnability is NP-hard.*

We prove this by a series of reductions. First, we prove a formal equivalence to the decision variant of k -incoherence (see Theorem 18). We then prove a reduction from optimization over k -incoherent matrices, and finally a reduction from this optimization problem to the k -CLIQUE problem for graphs, which is NP-hard.

Author note. Theorem 6 was also obtained independently and concurrently in [36]. We note that their work focused on k -incoherence, and they did not observe the equivalence to state classification proven in the present work.

1.3 Structure of the paper

In the next section, we establish a fundamental connection between the k -learnability of quantum states and the notion of k -incoherence of their Gram matrix. This provides the analytical

foundation for studying the classification problem. Section 3 formalizes both the decision and optimization variants of learnability by framing them as weak membership problems over a suitably defined convex set. We also introduce the weak optimization version of this problem and, in Section 4, present rigorous proofs showing that it can be used to solve the k -clique problem. The decision problem is shown to lie in NP in Section 5, through an analysis of the geometry of the set of k -incoherent matrices. Efficient algorithms for several special cases are then discussed in Section 6 and Section 7. Finally, we provide links to open-source software that implements the methods developed in this work.

2 How to tell if a set of states is k -learnable?

The main focus of this work is to study when a given set of states is k -learnable. We address this from both an analytical and computational complexity point of view.

To start, we model this as a semidefinite program (SDP). Suppose we are given an indexed list of states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ and a POVM $\{M_S : S \subseteq \{1, \dots, n\}, |S| = k\}$ which is intended to classify a given state into a subset of size k . Then, the average error this POVM makes is equal to the optimal objective function value of the following SDP

$$\min \left\{ \frac{1}{n} \sum_{i=1}^n \langle \psi_i | \left(\sum_{S: i \notin S} M_S \right) | \psi_i \rangle : M_S \in \text{Pos}(\mathbb{C}^d), \sum_{\substack{S \subseteq \{1, \dots, n\} \\ |S|=k}} M_S = I \right\} \quad (6)$$

where we denote the set of $n \times n$ positive semidefinite (resp. Hermitian) matrices as $\text{Pos}(\mathbb{C}^n)$ (resp. $\text{Herm}(\mathbb{C}^n)$). (Note that the minimum is attained since we are optimizing a continuous function over a compact set.) Thus, to determine if a list is k -learnable, one needs only to determine if the optimal value of the above SDP is 0 or not. However, special care must be taken when posing this question in the context of finite-precision algorithms.

When one examines the SDP above, there are a few things that are noteworthy when considering how difficult it is to solve. Often, one can efficiently approximate semidefinite programs with respect to their variable size and number of constraints, but the above SDP has $\binom{n}{k}$ matrix variables, each of size $n \times n$, which could be exponential when, e.g., $k \approx n/2$, as seen by Stirling's approximation. This suggests that this semidefinite program could be *hard* to approximate with respect to n and k in general.

2.1 Gram matrices and factor width

It turns out that determining whether states are k -learnable reduces to asking whether certain matrices can be decomposed as sums of positive semidefinite matrices with restricted support. To see this, consider the following SDP

$$\min \left\{ \sum_{i=1}^n \langle i | \sum_{S: i \notin S} W_S | i \rangle : W_S \in \text{Pos}(\mathbb{C}^n), \sum_{\substack{S \subseteq \{1, \dots, n\} \\ |S|=k}} W_S = \frac{1}{n} G \right\} \quad (7)$$

where G is the Gram matrix, i.e., $G_{i,j} = \langle \psi_i | \psi_j \rangle$.

This SDP ends up being useful to state the decision problem of classification, as seen in the following Lemma (proved in Appendix A).

Lemma 7. *Let k be a positive integer and let $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ be an indexed list of (possibly sub-normalized) pure quantum states. Then SDP (6) and SDP (7) have the same optimal value.*

Note that we are ultimately interested in whether either SDP has an optimal value 0. In this case, we have $\langle i|W_S|i\rangle = 0$ for each i and each S such that $i \notin S$. For such a pair (i, S) , this means that the (i, i) -th diagonal entry of W_S equals 0 and since W_S is positive semidefinite, the entire i -th row and column must be 0. Since this holds for all $i \notin S$, there must be $n - k$ rows and columns that are all 0. This aligns exactly with the notion of generalized incoherences, discussed next.

In the theory of quantum resources, k -incoherence [26, 37–39] plays the role of quantifying how much superposition is present in a (mixed) quantum state. This particular interpretation of it is not relevant for our purposes, so we just recall its mathematical definition.

Definition 8. *Let k be a positive integer. A positive semidefinite matrix X is called k -incoherent if there exists a positive integer m and a set of vectors v_i with the property that each v_i has at most k non-zero entries, such that*

$$X = \sum_{i=1}^m v_i v_i^*. \quad (8)$$

With this defined, we can now state the following, which is a generalization of a result from [20] from $n - 1$ to general k and defer the proof to Appendix B.

Lemma 9. *Let k be a positive integer. An indexed list of (possibly sub-normalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ is k -learnable if and only if its Gram matrix is k -incoherent.*

Another closely related definition is that of factor width [40–42], below.

Definition 10. *Let k be a positive integer. The factor width of a positive semidefinite matrix X is the smallest k such that it is k -incoherent.*

Equipped with this definition, we have the immediate corollary.

Corollary 11. *Let k be a positive integer. An indexed list of (possibly subnormalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ has learning width equal to the factor width of its Gram matrix.*

With this direct relationship between the learning width of pure states and the factor width of its Gram matrix, we now define the decision and optimization versions of these problems, and show that they are polynomial-time reducible to each other.

3 Weak membership and optimization problems

In order to discuss the computational complexity of k -incoherence, we consider it as a weak membership problem. We adopt the notation and definitions from [43] and apply it to the set of trace-bounded k -incoherent matrices, denoted by

$$\mathcal{I}_k^n := \{X \in \text{Herm}(\mathbb{C}^n) : \text{Tr}(X) \leq 1, X \text{ is } k\text{-incoherent}\}. \quad (9)$$

This is clearly a closed, convex set, and in Lemma 31 we prove that it has non-empty interior.

We define the following three sets of Hermitian matrices that are relevant to weak membership of sets with respect to the Frobenius norm defined as $\|A\|_F = \sqrt{\text{Tr}(A^* A)}$.

$$B(X, \delta) := \{Y : \|X - Y\|_F \leq \delta\} \quad (10)$$

$$S(\mathcal{I}_k^n, \delta) := \{X : \exists Y \in \mathcal{I}_k^n \text{ such that } \|X - Y\|_F \leq \delta\} \quad (11)$$

$$S(\mathcal{I}_k^n, -\delta) := \{X : B(X, \delta) \subseteq \mathcal{I}_k^n\} \quad (12)$$

The set (10) is the ball centered at X of radius δ , the set (11) is a δ -outer approximation of the set \mathcal{I}_k^n , and the set (12) is a δ -inner approximation of the set \mathcal{I}_k^n . Decision problems are typically formulated relative to these sets in order to avoid spurious computational difficulty associated with points close to the boundary.

We can now define the weak membership problem.

Definition 12 (WeakMEM(K, δ)). *The weak membership problem over a set K is as follows. Given a Hermitian matrix X and a real parameter $\delta > 0$, decide between the following two cases:*

- Yes case: $X \in S(K, -\delta)$;
- No case: $X \in S(K, \delta)$.

From this, we can define our first decision problem.

Definition 13 (FACTORWIDTH). *Given a positive integer k , $X \in \text{Herm}(\mathbb{C}^n)$, and real parameter $\delta > 0$, solve the weak membership problem WeakMEM(\mathcal{I}_k^n, δ).*

Before defining the weak membership problem for k -learnability, we take a closer look at the possible Gram matrices arising from a list of pure states. Consider the list $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ and its Gram matrix $G_{i,j} = \langle \psi_i | \psi_j \rangle$. For a normalized state, the corresponding diagonal entry is 1, and for subnormalized states it is strictly less than 1. We also note that Gram matrices are always positive semidefinite, and thus $\frac{1}{n}G$ is a trace-bounded positive semidefinite matrix. From Lemma 9, we know thus have $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ is k -learnable if and only if $\frac{1}{n}G \in \mathcal{I}_k^n$.

We thus arrive at the following membership problem for learning width.

Definition 14 (LEARNINGWIDTH). *Given k be a positive integer, an indexed list of (possibly subnormalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^n$, and a real parameter $\delta > 0$, solve the FACTORWIDTH problem for its normalized Gram matrix $\frac{1}{n}G$.*

Now that we have defined these two decision problems, we can state an immediate corollary.

Corollary 15. LEARNINGWIDTH and FACTORWIDTH are polynomial-time reducible to each other.

The proof follows since if given the states, one can compute the Gram matrix to precision δ and, conversely, if given a trace-bounded positive semidefinite matrix X , one can compute vectors $\{v_1, \dots, v_n\}$ such that $X_{i,j} = v_i^* v_j$ via the (efficient) Cholesky decomposition algorithm. In other words, there is a polynomial-time algorithm to map trace-bounded positive semidefinite operators to an indexed list of subnormalized pure quantum states, and vice versa.

We now define the optimization versions of the preceding decision problems.

Definition 16 (MINFACTORWIDTH). *Given a positive semidefinite matrix $X \in \text{Pos}(\mathbb{C}^n)$ of trace ≤ 1 , find the smallest k such that (k, X) is a YES instance to FACTORWIDTH.*

Definition 17 (MINLEARNINGWIDTH). *Given an indexed list of (possibly subnormalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^n$, find the smallest k such that $(k, \{|\psi_1\rangle, \dots, |\psi_n\rangle\})$ is a YES instance to LEARNINGWIDTH.*

By standard techniques, e.g., binary search, we have that the decision problem FACTORWIDTH and its optimization variant MINFACTORWIDTH are polynomial-time reducible to each other. Similarly, LEARNINGWIDTH and MINLEARNINGWIDTH are polynomial-time reducible to each other [44]. Combined with Corollary 15, we obtain the following.

Theorem 18. *The following problems are all polynomial-time reducible to each other:*

- LEARNINGWIDTH;
- MINLEARNINGWIDTH;
- FACTORWIDTH;
- MINFACTORWIDTH.

To show that LEARNINGWIDTH is NP-hard, and thus they all are, we examine the (weak) optimization over matrices with factor width at most k , next.

3.1 Weak optimization problems

In order to prove NP-hardness of testing weak membership in \mathcal{I}_k^n , we will use a reduction from *weak optimization* over \mathcal{I}_k^n , which we now define, following [43].

Definition 19 (WeakOPT(K, δ)). *The weak optimization problem over a set K is as follows. Given a Hermitian matrix C with $\|C\|_F = 1$, and real parameters γ and $\delta > 0$, decide between the following two cases:*

- Yes case: *There exists $X \in S(K, -\delta)$ with $\langle C, X \rangle \geq \gamma + \delta$;*
- No case: *For all $X \in S(K, \delta)$, we have $\langle C, X \rangle \leq \gamma - \delta$.*

The reduction from weak optimization to weak membership is given as follows.

Fact 20 (Theorem 2.3 in [43]). *Given a closed convex set $K \subseteq \text{Herm}(\mathbb{C}^n)$ satisfying*

$$B(X, r) \subseteq K \subseteq B(0, R), \quad (13)$$

for some $X \in \text{Herm}(\mathbb{C}^n)$, and real parameters r, R satisfying $R/r \leq \text{poly}(n)$. For any δ which is inverse-polynomial in n , there exists δ' which is inverse-polynomial in n for which there is a polynomial-time oracle reduction from WeakOPT(k, δ) to WeakMEM(k, δ').

For shorthand, we will use poly to denote an arbitrary polynomial function, and write this statement simply as WeakOPT($K, 1/\text{poly}$) reduces to WeakMEM($K, 1/\text{poly}$). We prove that this reduction holds for \mathcal{I}_k^n in the next section.

4 NP-hardness of FACTORWIDTH

In this section we prove that deciding weak membership in FACTORWIDTH is NP-hard.

Theorem 21. *WeakMEM($\mathcal{I}_k^n, 1/\text{poly}$) is NP-hard.*

Our starting point is a reduction from WeakOPT($\mathcal{I}_k^n, 1/\text{poly}$).

Lemma 22. *There is a polynomial-time oracle reduction from the problem WeakOPT($\mathcal{I}_k^n, 1/\text{poly}$) to the problem WeakMEM($\mathcal{I}_k^n, 1/\text{poly}$).*

This is a straightforward but slightly tedious application of Fact 20. See Appendix C. Hence, it suffices to prove NP hardness of $\text{WeakOPT}(\mathcal{I}_k^n, 1/\text{poly})$.

For a Hermitian matrix $C \in \text{Herm}(\mathbb{C}^n)$, consider the linear optimization problem

$$\mu(k, C) := \max\{\langle C, X \rangle : X \in \mathcal{I}_k^n\}. \quad (14)$$

(Note that the maximum is attained since we are optimizing a continuous function over a compact set.) If a linear optimization problem over a compact set has an optimal solution, then it has an optimal solution at an extreme point. We now characterize the extreme points of \mathcal{I}_k^n .

Lemma 23. *The set of extreme points of \mathcal{I}_k^n is*

$$P = \{vv^* : v \text{ has at most } k \text{ non-zero entries and } \|v\|_2 = 1\} \cup \{0\}. \quad (15)$$

This is proven in Appendix D. We can now succinctly write down the optimal value of the optimization problem $\mu(k, C)$. The following corollary is immediate.

Corollary 24. *For a Hermitian matrix $C \in \text{Herm}(\mathbb{C}^n)$, we have*

$$\mu(k, C) = \max_{S \subseteq [n], |S| \leq k} \lambda_{\max}(C_S), \quad (16)$$

where C_S the principal submatrix of C with rows and columns indexed by S , and λ_{\max} denotes the maximum eigenvalue.

In particular, if C is the adjacency matrix of a simple undirected graph G , then Corollary 24 shows that $\mu(k, C)$ is equal to the maximum eigenvalue of the adjacency matrix of an at-most- k -node subgraph. It can be shown that the maximum eigenvalue of a $k \times k$ adjacency matrix is maximized when it is a clique. This fact plays a critical component in our NP-hardness proof of $\text{WeakOPT}(\mathcal{I}_k^n, 1/\text{poly})$.

Definition 25 (CLIQUE [45]). *Given a simple undirected nonempty graph G and a positive integer k , decide whether G contains a clique of size k .*

It is known that CLIQUE is NP-hard [45]. The following proposition shows that CLIQUE reduces to $\text{WeakOPT}(\mathcal{I}_k^n, 1/\text{poly})$ for \mathcal{I}_k^n , completing the proof of Theorem 21.

Proposition 26. *Let $n \geq k \geq 2$ be integers, let C be the adjacency matrix of a simple nonempty undirected graph G on n vertices with e edges, and let $C' = \frac{1}{\sqrt{2e}}C$ be the normalization in Frobenius norm. Then there exist functions $\gamma = \gamma(e, k, n) \in \mathbb{R}$ and $\delta = 1/\text{poly}(n)$ for which*

- $\mu(k, C') \geq \gamma + \delta$ if G contains a k -clique.
- $\mu(k, C') \leq \gamma - \delta$ if G does not contain a k -clique.

See Appendix E for the proof. The key idea is to show that $\mu(k, C) = k - 1$ if G contains a k -clique, and $\mu(k, C) \leq k - 1 - \frac{1}{\text{poly}(k)}$ if G does not contain a k -clique.

5 Succinct proofs of FACTORWIDTH

In this section, we show that weak membership is in NP, i.e., $\text{WeakMEM}(\mathcal{I}_k^n, 1/\text{poly}) \in \text{NP}$, by showing how to find *succinct* proofs. For this discussion, it helps to recall the definition of \mathcal{I}_k^n :

$$\mathcal{I}_k^n = \{X : X = \sum_i v_i v_i^*, \text{ where each } v_i \text{ has at most } k \text{ non-zero entries, } \text{Tr}(X) \leq 1\}. \quad (17)$$

A first thought to prove that $X \in \mathcal{I}_k^n$ is to simply provide all of the v_i vectors and check the conditions above. However, there are an exponential number of these vectors (at least $\binom{n}{k}$), making this proof too long (it must be polynomial in size).

The trick to reducing the dimension is notice that $\text{Herm}(\mathbb{C}^n)$ is a vector space of dimension n^2 , so Carathéodory's theorem (see, e.g., [1, Theorem 1.9]) tells us that any member of \mathcal{I}_k^n can be written as a convex combination of $n^2 + 1$, or fewer, extreme points of that convex set. Since we have already characterized the extreme points in Lemma 23, we know that can write any $X \in \mathcal{I}_k^n$ as

$$X = \sum_{i=1}^{n^2+1} v_i v_i^* \quad (18)$$

where v_i has at most k non-zero entries. (We absorbed the convex combination scalars into the definitions of v_i .) Thus, a short proof of membership in \mathcal{I}_k^n simply provides the v_i matrices above, checking that they have the proper support, Equation (18), and its trace.

We remark that even though this proof is succinct, this does not imply it is easy to find. Indeed, when optimizing or testing membership of this set, we do not know which polynomially many of the exponentially many v_i s are the ones we can use.

6 Polynomial-time algorithm for FACTORWIDTH for constant k

Our approach to efficiently test whether an indexed list of (possibly subnormalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ is k -learnable is to solve the weak membership problem over \mathcal{I}_k^n of its (scaled) Gram matrix $\frac{1}{n}G$.

Before continuing, we prove a lemma to bound the Frobenius norm of any matrix we are considering to be in $S(\mathcal{I}_k^n, \delta)$.

Lemma 27. *Given X and δ , if $\|X\|_F > 1 + \delta$, then $X \notin S(\mathcal{I}_k^n, \delta)$.*

Proof. By the reverse triangle inequality, we have

$$\|X - Y\|_F \geq \|X\|_F - \|Y\|_F > \delta \quad (19)$$

for any $Y \in \mathcal{I}_k^n$. □

Therefore, we have a two-step algorithm. First, check if $\|X\|_F > 2 + \delta$. If so, return “No”, $X \notin S(\mathcal{I}_k^n, \delta)$. Otherwise, solve an SDP which we describe below. Given X , the matrix we wish to test \mathcal{I}_k^n membership of, one can solve the following semidefinite program

$$\alpha = \min\{\|X - Y\|_F^2 : Y \in \mathcal{I}_k^n\} \quad (20)$$

to within precision $\delta^2/3$. After doing this, it is clear whether there exists $Y \in \mathcal{I}_k^n$ δ -close to X , and thus to respond with “Yes” or “No”.

Before continuing, we present the well-known semidefinite programming formulation of the Frobenius norm, for completeness.

Given X (not necessarily square), consider the block matrix

$$M = \begin{pmatrix} I & X \\ X^* & Z \end{pmatrix}. \quad (21)$$

Then by looking at Schur complements, we have M positive semidefinite if and only if $Z - X^*X$ is positive semidefinite. Thus, minimizing the trace of Z yields $\text{Tr}(X^*X) = \|X\|_F^2$.

Thus, we can write SDP (20) as the SDP below.

$$\alpha = \min \left\{ \text{Tr}(Z) : \begin{pmatrix} I & Y - X \\ Y - X & Z \end{pmatrix} \geq 0, Y \in \mathcal{I}_k^n \right\} \quad (22)$$

$$= \min \left\{ \text{Tr}(Z) : \begin{pmatrix} I & Y - X \\ Y - X & Z \end{pmatrix} \geq 0, \sum_{S:|S|=k} W_S = Y, W_S \geq 0, \langle i | W_S | i \rangle = 0, \forall i \notin S, \text{Tr}(Y) \leq 1 \right\}. \quad (23)$$

It is helpful to bound the trace of Z without affecting the value of α (so it does not cut off any near optimal solutions). We know that in any optimal solution we have

$$\text{Tr}(Z) = \|X - Y\|_F^2 \leq \|X\|_F^2 \leq (\delta + 2)^2. \quad (24)$$

where the first inequality holds since $Y = 0$ is a feasible solution.

Thus, the SDP we actually solve takes the form

$$\alpha = \min \left\{ \text{Tr}(Z) : \begin{pmatrix} I & Y - X \\ Y - X & Z \end{pmatrix} \geq 0, \sum_{S:|S|=k} W_S = Y, W_S \geq 0, \langle i | W_S | i \rangle = 0, \forall i \notin S, \right. \quad (25)$$

$$\left. \text{Tr}(Y) \leq 1, \text{Tr}(Z) \leq (\delta + 2)^2 \right\}. \quad (26)$$

Assuming we can write the SDP in standard form

$$\alpha = \sup \{ \langle A, X \rangle : \Phi(X) = B, X \geq 0 \}, \quad (27)$$

then one can approximate α to $\epsilon > 0$ precision (additive error) in time $\text{poly}(a, b, \log(R), \log(1/\epsilon), c)$ using the Ellipsoid Algorithm [35], where:

- X is $a \times a$;
- B is $b \times b$;
- $\|X\|_2 \leq R$ for all feasible X (and there exists feasible X);
- c is the maximum bit-length of the entries in A, B , and $J(\Phi)$, where J is the Choi representation of Φ .

To modify our SDP, we first write the linear operator

$$\Psi_1(Y, Z, W) = \begin{pmatrix} 0 & Y \\ Y & Z \end{pmatrix} - W \quad \text{and} \quad B_1 = \begin{pmatrix} -I & X \\ X & 0 \end{pmatrix} \quad (28)$$

where Y, Z, W are all positive semidefinite. Then $\Psi_2(Y, Z, W) = B_1$ models the

$$\begin{pmatrix} I & Y - X \\ Y - X & Z \end{pmatrix} \geq 0 \quad (29)$$

constraint. If we define

$$\Psi_2(Y, W_1, \dots, W_{\binom{n}{k}}) = \sum_{i=1}^{\binom{n}{k}} W_i - Y \quad \text{and} \quad B_2 = 0 \quad (30)$$

this models the $Y = \sum_{S:|S|=k} W_S$ constraint. If we define

$$\Phi_S(W_S) = \sum_{i \notin S} \langle i | W_S | i \rangle \quad \text{and} \quad \tilde{B}_S = 0 \quad (31)$$

this models the $\langle i | W_S | i \rangle = 0$ constraints. If we define

$$\Psi_3(Y, t) = \langle I, Y \rangle + t \quad \text{and} \quad B_3 = 1 \quad (32)$$

where $t \geq 0$ this models the $\text{Tr}(Y) \leq 1$ constraint. If we define

$$\Psi_4(Z, s) = \langle I, Z \rangle + s \quad \text{and} \quad B_4 = (\delta + 2)^2 \quad (33)$$

for $s \geq 0$ this models the $\text{Tr}(Z) \leq (\delta + 2)^2$ constraint.

Therefore, we have the variables $(Y, Z, W, W_1, \dots, W_{\binom{n}{k}}, t, s)$, all positive semidefinite. Also, we have the constants $B_1, B_2, B_3, B_4, \tilde{B}_1, \dots, \tilde{B}_{\binom{n}{k}}$. By defining a single SDP variable as

$$Y \oplus Z \oplus W \oplus W_1 \oplus t \oplus s \oplus W_1 \oplus \dots \oplus W_{\binom{n}{k}} \quad (34)$$

and a single constant matrix as

$$B_1 \oplus B_2 \oplus B_3 \oplus B_4 \oplus \tilde{B}_1 \oplus \dots \oplus \tilde{B}_{\binom{n}{k}} \quad (35)$$

and a single linear map Ψ as

$$\Psi(Y, Z, W, W_1, \dots, W_{\binom{n}{k}}, t, s) = \Psi_1 \oplus \Psi_2 \oplus \Psi_3 \oplus \Psi_4 \oplus \Phi_1 \oplus \dots \oplus \Phi_{\binom{n}{k}} \quad (36)$$

we can write the affine SDP constraints as $\Psi(X) = B$. Lastly, define A as

$$0 \oplus I \oplus 0 \oplus 0 \oplus 0 \oplus \dots \oplus 0 \oplus 0 \quad (37)$$

(where the dimensions of the 0 matrices align with Equation (34)) finishes putting the SDP in standard form.

First notice that the bit-length of the data of the standard form SDP is the same as X and δ .

It is now straightforward to check that we can take

- $R = \text{poly}(n, \delta) = \text{poly}(n)$ when $\delta = 1/\text{poly}(n)$.
- $a = \text{poly}(n)$ when k is constant.
- $b = \text{poly}(n)$.
- $\epsilon = \delta^2/3$ (this suffices to tell if there is a $Y \in \mathcal{I}_k^n$ such that $\|X - Y\|_F \leq \delta$ or not).

Thus, the SDP runs in time $\text{poly}(n)$ and the bit-length of X and δ .

7 Polynomial-time algorithm for FACTORWIDTH when the rank is constant

In this section, we give a $\text{poly}(n)$ -time algorithm for computing the factor width of a matrix $M \in \text{Pos}(\mathbb{C}^n)$ when the rank of M is fixed (and k is arbitrary). Note that this fact is trivial in the rank 1 case (and is exploited considerably in the rank 1 case in [36]).

Theorem 28. *Let $M \in \text{Pos}(\mathbb{C}^n)$ has rank r , and k is a positive integer. Then there exists a positive integer $m \in O(n^r)$ and subspaces $T_1, \dots, T_m \subseteq \mathbb{C}^n$ (which can be found in $\text{poly}(n)$ time) with the property that M is k -incoherent if and only if there exist matrices $M_1, \dots, M_m \in \text{Pos}(\mathbb{C}^n)$ for which*

$$M = \sum_{j=1}^m M_j \quad \text{and} \quad M_j = \Pi_j M_j \Pi_j \quad (38)$$

where, for all $j \in \{1, 2, \dots, m\}$, Π_j is the orthogonal projection onto T_j .

We note that determining the existence of M_1, \dots, M_m can be solved by semidefinite programming, and approximated in a similar way as was done in Section 6.

Before proving this theorem, we note that it immediately implies a $\text{poly}(n)$ -time algorithm for determining k -learnability of n pure states in any constant dimension d . This follows from the fact that the rank of a Gram matrix is exactly the minimum dimension of states $|\psi_1\rangle, \dots, |\psi_n\rangle$ with that Gram matrix.

Corollary 29. *Suppose $S = \{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ and d is constant. Then the learning width of S can be determined in $\text{poly}(n)$ time by solving a polynomial number of polynomial-sized linear systems, and a polynomial-sized semidefinite program.*

Now we prove Theorem 28.

Proof of Theorem 28. Let $S := \text{range}(M) \subseteq \mathbb{C}^n$, so $\dim(S) = r$. For each $j \in \{1, 2, \dots, n\}$, consider the subspace

$$S_j := S \cap \{v \in \mathbb{C}^n : v(j) = 0\}. \quad (39)$$

Similarly, for an index set $J \subseteq \{1, 2, \dots, n\}$, define

$$S_J := \bigcap_{j \in J} S_j = S \cap \{v \in \mathbb{C}^n : v(j) = 0 \ \forall j \in J\}. \quad (40)$$

Let $\{T_1, T_2, \dots, T_m\}$ be the set of such subspaces corresponding to all $J \subseteq \{1, 2, \dots, n\}$ with $|J| = n - k$. Our first goal is to show that $m \leq \sum_{j=0}^r \binom{n}{j} \in O(n^r)$ and that these subspaces can be found in polynomial time.

To this end, we proceed recursively. Start with the set $R_0 := \{S\}$. Then, for each integer $j \in \{1, 2, \dots, n - k\}$, we define

$$R_j := \{R \cap \{v \in \mathbb{C}^n : v(i) = 0\} : R \in R_{j-1}, i \in \{1, 2, \dots, n\}\}. \quad (41)$$

For each $R \in R_{j-1}$ and $i \in \{1, 2, \dots, n\}$, we have either $R \cap \{v \in \mathbb{C}^n : v(i) = 0\} = R$ or $\dim(R \cap \{v \in \mathbb{C}^n : v(i) = 0\}) = \dim(R) - 1$. Since $\dim(S) = r$, this recursive construction cannot decrease the dimension of S more than r times, so the total number of distinct subspaces in R_j can never exceed $\sum_{j=0}^r \binom{n}{j}$. Since $R_{n-k} = \{T_1, T_2, \dots, T_m\}$, the fact that $m \leq \sum_{j=0}^r \binom{n}{j}$ follows.

To see that (orthonormal bases of) T_1, T_2, \dots, T_m can be computed in $\text{poly}(n)$ time, simply notice that (as noted earlier) each R_{j-1} contains at most $\sum_{j=0}^r \binom{n}{j}$ subspaces, so each R_j can be constructed by computing orthonormal bases for at most $n \sum_{j=0}^r \binom{n}{j}$ subspaces. Since the recursion runs for $n - k$ steps, this amounts to computing at most $(n - k)n \sum_{j=0}^r \binom{n}{j} \leq n^2 \sum_{j=0}^r \binom{n}{j}$ subspaces, which is polynomial in n .

Our next goal is to show that M is k -incoherent if and only if there exist matrices $M_1, \dots, M_m \in \text{Pos}(\mathbb{C}^n)$ for which (38) holds. To this end, note that if M can be written in this form then it has factor width at most k , since every vector in T_j has at most k non-zero entries, so each $M_j = \Pi_j M_j \Pi_j$ equals 0 outside of a single $k \times k$ principal submatrix. Conversely, if M cannot be written in this form then, since $\bigcup_{j=1}^m T_j$ contains every vector $v \in S$ with at most k non-zero entries, M must have factor width strictly larger than k . \square

Example 30. To illustrate how Theorem 28 works, suppose we want to determine the factor width of the rank-3 matrix

$$M = \begin{bmatrix} 2 & 1 & 1 & -1 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & -1 \\ -1 & 1 & -1 & 2 \end{bmatrix}. \quad (42)$$

We start by finding a basis for $S := \text{range}(M)$, which can be done by picking a linearly independent set of $r = 3$ columns of M : $S = \text{span}\{(2, 1, 1, -1), (1, 2, 0, 1), (1, 0, 2, -1)\}$. Then $R_0 = \{S\}$ and we proceed recursively as in the proof of Theorem 28:

$$\begin{aligned} R_1 = \{S_1, S_2, S_3, S_4\}, \quad \text{where} \quad & S_1 = \text{span}\{(0, 1, -1, 1), (0, 1, -3, 1)\}, \\ & S_2 = \text{span}\{(1, 0, 2, -1), (3, 0, 2, -3)\}, \\ & S_3 = \text{span}\{(1, 2, 0, 1), (3, 2, 0, -1)\}, \text{ and} \\ & S_4 = \text{span}\{(1, 1, 1, 0), (3, 3, 1, 0)\}. \end{aligned}$$

To determine whether or not M is 3-incoherent, we let Π_1, Π_2, Π_3 , and Π_4 be the orthogonal projections onto S_1, S_2, S_3 , and S_4 , respectively. We then use semidefinite programming to determine whether or not there exist matrices $M_1, M_2, M_3, M_4 \in \text{Pos}(\mathbb{C}^4)$ for which

$$M = M_1 + M_2 + M_3 + M_4, \quad \text{and} \quad M_j = \Pi_j M_j \Pi_j \quad \text{for all } j \in \{1, 2, 3, 4\}. \quad (43)$$

Indeed, such matrices do exist:

$$M_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad M_4 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (44)$$

so M is 3-incoherent.

To similarly determine whether or not M is 2-incoherent, we proceed further with the recursive construction by computing

$$\begin{aligned} R_2 = \{S_{\{1,2\}}, S_{\{1,3\}}, S_{\{2,3\}}, S_{\{3,4\}}\}, \quad \text{where} \quad & S_{\{1,2\}} = S_{\{1,4\}} = S_{\{2,4\}} = \text{span}\{(0, 0, 1, 0)\}, \\ & S_{\{1,3\}} = \text{span}\{(0, 1, 0, 1)\}, \\ & S_{\{2,3\}} = \text{span}\{(1, 0, 0, -1)\}, \text{ and} \\ & S_{\{3,4\}} = \text{span}\{(1, 1, 0, 0)\}. \end{aligned}$$

It follows that the only vectors in $\text{range}(M)$ with $k = 2$ or fewer non-zero entries are the scalar multiples of $v_1 := (0, 0, 1, 0)$, $v_2 := (0, 1, 0, 1)$, $v_3 := (1, 0, 0, -1)$, and $v_4 := (1, 1, 0, 0)$, so M is 2-incoherent if and only if there exist non-negative real scalars c_1 , c_2 , c_3 , and c_4 for which

$$M = c_1 v_1 v_1^* + c_2 v_2 v_2^* + c_3 v_3 v_3^* + c_4 v_4 v_4^*. \quad (45)$$

It is straightforward to use semidefinite programming (or even just solve by hand in this small example) to see that no such scalars exist, so X is not 2-incoherent. It follows that X has factor width 3.

Software

Companion software implements numerical routines for determining k -incoherence of matrices, the factor width, and k -learnability of quantum state collections. These are provided within the `toqito` quantum information package [46], leveraging `PICOS` [47] and the `CVXOPT` solver [48] for semidefinite program optimization.

Acknowledgements

The authors thank Angus Lowe and Richard Cleve for helpful conversations. N.J. was supported by NSERC Discovery Grant number RGPIN-2022-04098. J.S. is funded in part by the Commonwealth of Virginia's Commonwealth Cyber Initiative (CCI) under grant number 469351.

References

- [1] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.
- [2] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [3] Mark M Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.
- [4] Carl W Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, 1969.
- [5] Viacheslav P Belavkin. Optimal distinction of non-orthogonal quantum signals. *Radio Engineering and Electronic Physics*, 20:39–47, 1975.
- [6] Viacheslav P Belavkin. Optimal multiple quantum statistical hypothesis testing. *Stochastics: An International Journal of Probability and Stochastic Processes*, 1(1-4):315–345, 1975.
- [7] Matthieu E Deconinck and Barbara M Terhal. Qubit state discrimination. *Physical Review A*, 81(6):062304, 2010.
- [8] Shashank Virmani, Massimiliano F Sacchi, Martin B Plenio, and Damian Markham. Optimal local discrimination of two multipartite pure states. *Physics Letters A*, 288(2):62–68, 2001.
- [9] Jonathan Walgate, Anthony J Short, Lucien Hardy, and Vlatko Vedral. Local distinguishability of multipartite orthogonal quantum states. *Physical Review Letters*, 85(23):4972, 2000.
- [10] Joonwoo Bae and Leong-Chuan Kwek. Quantum state discrimination and its applications. *Journal of Physics A: Mathematical and Theoretical*, 48(8):083001, 2015.

- [11] Stephen M. Barnett and Sarah Croke. Quantum state discrimination. *Advances in Optics and Photonics*, 1(2):238–278, 2009.
- [12] Anthony Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.
- [13] János A. Bergou. Discrimination of quantum states. *Journal of Modern Optics*, 57(3):160–180, 2010.
- [14] Somshubhro Bandyopadhyay, Rahul Jain, Jonathan Oppenheim, and Christopher Perry. Conclusive exclusion of quantum states. *Physical Review A*, 89(2):022336, 2014.
- [15] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Conditions for compatibility of quantum-state assignments. *Physical Review A*, 66(6):062111, 2002.
- [16] Vojtěch Havlíček and Jonathan Barrett. Simple communication complexity separation from quantum state antidistinguishability. *Physical Review Research*, 2(1):013326, 2020.
- [17] Teiko Heinosaari and Oskari Kerppo. Antidistinguishability of pure quantum states. *Journal of Physics A: Mathematical and Theoretical*, 51(36):365303, 2018.
- [18] Hemant K. Mishra, Michael Nussbaum, and Mark M. Wilde. On the optimal error exponents for classical and quantum antidistinguishability. *arXiv preprint arXiv:2309.03723*, 2023.
- [19] Vincent Russo and Jamie Sikora. Inner products of pure states and their antidistinguishability. *Physical Review A*, 107:L030202, 2023.
- [20] Nathaniel Johnston, Vincent Russo, and Jamie Sikora. Tight bounds for antidistinguishability and circulant sets of pure quantum states. *Quantum*, 9:1622, 2025.
- [21] Matthew Saul Leifer. Is the quantum state real? An extended review of ψ -ontology theorems. *arXiv preprint arXiv:1409.1570*, 2014.
- [22] Matthew Leifer and Cristhiano Duarte. Noncontextuality inequalities from antidistinguishability. *Physical Review A*, 101:062113, 2020.
- [23] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, 2012.
- [24] Abel Molina. POVMs are equivalent to projections for perfect state exclusion of three pure states in three dimensions. *Quantum*, 3:117, 2019.
- [25] Jonathan Crickmore, Ittoop V Puthoor, Berke Ricketti, Sarah Croke, Mark Hillery, and Erika Andersson. Unambiguous quantum state elimination for qubit sequences. *Physical Review Research*, 2(1):013256, 2020.
- [26] Martin Ringbauer, Thomas R Bromley, Marco Cianciaruso, Ludovico Lami, WY Sarah Lau, Gerardo Adesso, Andrew G White, Alessandro Fedrizzi, and Marco Piani. Certification and quantification of multilevel quantum coherence. *Physical Review X*, 8(4):041007, 2018.
- [27] Maria Schuld and Francesco Petruccione. *Machine Learning with Quantum Computers*, volume 676. Springer, 2021.
- [28] Peter Wittek. *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. Academic Press, 2014.

- [29] Srinivasan Arunachalam and Ronald De Wolf. Guest column: A survey of quantum learning theory. *ACM Sigact News*, 48(2):41–67, 2017.
- [30] Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Reviews Physics*, 6(1):59–69, 2024.
- [31] Daniel FV James, Paul G Kwiat, William J Munro, and Andrew G White. Measurement of qubits. *Physical Review A*, 64(5):052312, 2001.
- [32] 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.
- [33] Alexander Semenovovich Holevo. Information-theoretical aspects of quantum measurement. *Problemy Peredachi Informatsii*, 9(2):31–42, 1973.
- [34] Asher Peres and William K Wootters. Optimal detection of quantum information. *Physical Review Letters*, 66(9):1119, 1991.
- [35] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.
- [36] Aram W Harrow, Angus Lowe, and Freek Witteveen. Randomized truncation of quantum states. *arXiv preprint arXiv:2510.08518*, 2025.
- [37] 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*, 106:052417, 2022.
- [38] Federico Levi and Florian Mintert. A quantitative theory of coherent delocalization. *New Journal of Physics*, 16(3):033007, 2014.
- [39] Jan Sperling and Werner Vogel. Convex ordering and quantification of quantumness. *Physica Scripta*, 90(7):074024, 2015.
- [40] Francesco Barioli and Abraham Berman. The maximal CP-rank of rank k completely positive matrices. *Linear Algebra and its Applications*, 363:17–33, 2003.
- [41] Nathaniel Johnston, Shirin Moein, and Sarah Plosker. The factor width rank of a matrix. *Linear Algebra and its Applications*, 716:32–59, 2025.
- [42] 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.
- [43] Yi-Kai Liu. *The complexity of the consistency and N -representability problems for quantum states*. University of California, San Diego, 2007.
- [44] Jon Kleinberg and Eva Tardos. *Algorithm Design*. Pearson Education India, 2006.
- [45] Richard M Karp. Reducibility among combinatorial problems. In *50 Years of Integer Programming 1958-2008: from the Early Years to the State-of-the-Art*, pages 219–241. Springer, 2009.
- [46] Vincent Russo. toqito – Theory of quantum information toolkit: A Python package for studying quantum information. *Journal of Open Source Software*, 6(61):3082, 2021.

- [47] Guillaume Sagnol and Maximilian Stahlberg. PICOS: A Python interface to conic optimization solvers. *Journal of Open Source Software*, 7(70):3915, 2022.
- [48] Martin Andersen, Joachim Dahl, and Lieven Vandenbergh. CVXOPT: Convex Optimization. *Astrophysics Source Code Library*, page ascl:2008.017, 2020. ADS Bibcode: 2020ascl.soft08017A.
- [49] 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:2312.04023*, 2023.
- [50] Suman Chakraborty ([https://math.stackexchange.com/users/885456/suman chakraborty](https://math.stackexchange.com/users/885456/suman%20chakraborty)). Upper bound on the largest eigenvalue of a simple graph. Mathematics Stack Exchange. URL:<https://math.stackexchange.com/q/4138764> (version: 2021-05-14).

A Proof of Lemma 7

We reproduce SDP (6), below

$$\alpha = \min \left\{ \frac{1}{n} \sum_{i=1}^n \langle \psi_i | \left(\sum_{S: i \notin S} M_S \right) | \psi_i \rangle \right\} \quad (46)$$

where the minimization is over POVMS.

Define $r_i = \|\psi_i\|_2$ and

$$|\phi_i\rangle = \begin{cases} (1/r_i)|\psi_i\rangle & \text{when } r_i \neq 0 \\ \text{To be determined} & \text{when } r_i = 0 \end{cases} \quad (47)$$

to make $|\psi_i\rangle = r_i|\phi_i\rangle$. Notice that we have the freedom to choose states above, and we do this such that each one is orthogonal to everything else. We can do this by increasing the dimension if need be and this particular choice will ultimately make the corresponding Gram matrix nice. For an index i and subset S , define

$$t_{i,S} = \begin{cases} 0 & \text{when } i \in S \\ 1 & \text{when } i \notin S. \end{cases} \quad (48)$$

Lastly, define $q_i = 1/n$ and $R_{i,S} = r_i^2 t_{i,S}$, for all i and S .

With these things defined, we can write the SDP in the following way

$$\alpha = \min \left\{ \sum_{i=1}^n \sum_{S: |S|=k} q_i R_{i,S} \langle M_S, |\psi_i\rangle \langle \psi_i| \rangle \right\} \quad (49)$$

where the minimization is over POVMS (the explicit constraints are omitted for clarity). When written in this way, the objective function can be interpreted as the average *reward* when given $|\psi_i\rangle$ with probability q_i , and measured with the POVM $(M_S : |S| = k)$ to receive reward $R_{i,j}$. This framework has been studied in [49] and it can be shown that the optimal value of the preceding SDP equals that of the SDP

$$\alpha = \min \left\{ \sum_{i=1}^n \sum_{S: |S|=k} q_i R_{i,S} \langle W_S, |i\rangle \langle i| \rangle : \sum_{S: |S|=k} W_S = \tilde{G}, W_S \geq 0 \right\} \quad (50)$$

where we used \tilde{G} as the Gram matrix for the (normalized) states $\{|\phi_1\rangle, \dots, |\phi_n\rangle\}$.

By substituting most of the parameters back in, we get

$$\alpha = \min \left\{ \frac{1}{n} \sum_{i=1}^n r_i^2 \langle i | \left(\sum_{i \notin S} W_S \right) | i \rangle : \sum_{S:|S|=k} W_S = \tilde{G}, W_S \geq 0 \right\} \quad (51)$$

$$= \min \left\{ \sum_{i=1}^n r_i^2 \langle i | \left(\sum_{i \notin S} \tilde{W}_S \right) | i \rangle : \sum_{S:|S|=k} \tilde{W}_S = \frac{1}{n} \tilde{G}, \tilde{W}_S \geq 0 \right\} \quad (52)$$

The rest of the analysis is easier if we define a diagonal matrix and its (generalized) inverse. If we define the diagonal matrix $D_{i,i} = r_i$, then we can rewrite $r_i |i\rangle = D|i\rangle$ and also the Gram matrix G of the original states as

$$G_{i,j} = \langle \psi_i | \psi_j \rangle = r_i r_j \langle \phi_i | \phi_j \rangle = (D \tilde{G} D^*)_{i,j}. \quad (53)$$

Define the Moore-Penrose generalized inverse of D as D^+ where $(D^+)_{i,j} = 1/r_i$ when $r_i \neq 0$, and 0 when $r_i = 0$. Then we have, for i, j such that $r_i, r_j \neq 0$:

$$\tilde{G}_{i,j} = \langle \phi_i | \phi_j \rangle = \frac{1}{r_i r_j} \langle \psi_i | \psi_j \rangle = (D^+ G D^+)_{i,j}. \quad (54)$$

From this, it is easy to see by the choice of $|\phi_i\rangle$ when $r_i = 0$ that

$$\tilde{G} = D^+ G D^+ + I_T, \quad (55)$$

where T only has support on the entries where $r_i = 0$.

With this, the SDP reduces to

$$\alpha = \min \left\{ \sum_{i=1}^n \langle i | \left(\sum_{i \notin S} D \tilde{W}_S D \right) | i \rangle : \sum_{S:|S|=k} \tilde{W}_S = \frac{1}{n} \tilde{G}, \tilde{W}_S \geq 0 \right\}. \quad (56)$$

To finish the proof, we now argue that α is also equal to the optimal objective function value of SDP (7) below, which we have defined as β .

$$\beta = \min \left\{ \sum_{i=1}^n \langle i | \sum_{S: i \notin S} W_S | i \rangle : \sum_{S:|S|=k} W_S = \frac{1}{n} G, W_S \geq 0 \right\} \quad (57)$$

Proof that $\beta \leq \alpha$. Suppose $(\tilde{W}_S : |S| = k)$ is feasible in the α SDP (56) with objective function value t . Define the following $W_S := D \tilde{W}_S D^* \geq 0$. Since $(\tilde{W}_S : |S| = k)$ is feasible, we know that $\sum_{S:|S|=k} \tilde{W}_S = (1/n) \tilde{G}$, and thus

$$\sum_{S:|S|=k} W_S = \sum_{S:|S|=k} D \tilde{W}_S D^* = (1/n) D \tilde{G} D^* = (1/n) G. \quad (58)$$

Thus $(W_S : |S| = k)$ is feasible in the β SDP (57) with objective function value

$$\sum_{i=1}^n \langle i | \sum_{S: i \notin S} W_S | i \rangle = \sum_{i=1}^n \langle i | \sum_{S: i \notin S} D \tilde{W}_S D | i \rangle = t. \quad (59)$$

Thus, $\beta \leq \alpha$.

Proof that $\alpha \leq \beta$. Let $(W_S : |S| = k)$ be feasible in the β SDP (57) with objective function value s . Define $\tilde{W}_S := D^+ W_S D^+ \geq 0$. Since $(W_S : |S| = k)$ is feasible, we have that $\sum_{S:|S|=k} W_S = (1/n)G$. Thus we have

$$\sum_{S:|S|=k} \tilde{W}_S = \sum_{S:|S|=k} D^+ W_S D^+ = \sum_{S:|S|=k} (1/n) D^+ G D^+ \quad (60)$$

which is *almost feasible*. If we add $(1/n)I_T$ to any of the W_S , it will still be positive semidefinite and now $(\tilde{W}_S : |S| = k)$ is feasible in the α SDP (56). Note that this addition of I_T does not contribute to the objective function value since $D I_T D = 0$, so we are not concerned to which variable it is added.

All that remains is to compute the objective function value of this feasible solution. Notice that since $G_{i,i} = 0$ for any i such that $r_i = 0$, it is also the case that $(W_S)_{i,i} = 0$ for all S . We can now compute the objective function value as

$$\sum_{i=1}^n \langle i | \sum_{S:i \notin S} D \tilde{W}_S D | i \rangle = \sum_{i=1}^n \langle i | \sum_{S:i \notin S} D D^+ W_S D^+ D | i \rangle = \sum_{i:r_i \neq 0} \langle i | \sum_{S:i \notin S} W_S | i \rangle = \sum_i \langle i | \sum_{S:i \notin S} W_S | i \rangle = s. \quad (61)$$

Thus, we have $\alpha \leq \beta$ as required.

B Proof of Lemma 9

Here we prove Lemma 9 which states that an indexed list of (possibly sub-normalized) pure quantum states $\{|\psi_1\rangle, \dots, |\psi_n\rangle\} \subset \mathbb{C}^d$ is k -learnable if and only if its Gram matrix is k -incoherent.

“If” direction. Let G be the Gram matrix and suppose it is k -incoherent. Then SDP (7) has objective value 0, and by Lemma 7, so does SDP (6). Since that SDP attains an optimal solution as the feasible region is the set of valid POVMs with $\binom{n}{k}$ outcomes, and thus compact, we know the existence of a POVM which satisfies the conditions for k -learnability.

“Only if” direction. Suppose the set is k -learnable. Then SDP (6) has objective function 0, and by Lemma 7, so does SDP (7). If SDP (7) attains an optimal solution, then said solution asserts that G is k -incoherent. Thus, showing attainment of an optimal solution suffices to finish the proof.

We proceed with a similar argument as above, that the feasible region is bounded, and thus compact. Since each W_S is positive semidefinite, we have that $\|W_S\|_{\text{Tr}} = \text{Tr}(W_S) \leq \frac{1}{n} \text{Tr}(G) \leq 1$. Thus, the feasible region is bounded in the trace-norm (and hence any norm since we are working in finite dimensions).

C Proof of Lemma 22

Here we prove Lemma 22, which states that the problem $\text{WeakOPT}(\mathcal{I}_k^n, 1/\text{poly})$ reduces to the problem $\text{WeakMEM}(\mathcal{I}_k^n, 1/\text{poly})$. We require a lemma and a corollary to prove this result.

Lemma 31. *Let $M \in \text{Herm}(\mathbb{C}^n)$. If $\|M - \frac{I}{n}\|_{\text{F}} \leq 1/d$ then M is a 2-incoherent positive operator (and thus k -incoherent for all $k \in \{2, \dots, n\}$).*

In particular, this lemma demonstrates that \mathcal{I}_k^n has non-empty interior.

Proof. Multiplying by n , it suffices to prove that if $\|M - I\|_F \leq 1$, then M is 2-incoherent. It is easy to see that M must be positive semidefinite under this condition. Observe that

$$\|M - I\|_F^2 = \sum_{j=1}^n (\lambda_j - 1)^2 = n - 2 \operatorname{Tr}(M) + \sum_{j=1}^n \lambda_j^2. \quad (62)$$

We know from [37, Theorem 7] that if $M \in \operatorname{Pos}(\mathbb{C}^n)$ has $\operatorname{Tr}(M) = 1$ and eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ then $\sum_{j=1}^n \lambda_j^2 \leq 1/(n-1)$ implies that M is 2-incoherent. By rescaling M , we see that this is equivalent (for all $M \in \operatorname{Pos}(\mathbb{C}^n)$, not just those with $\operatorname{Tr}(M) = 1$) to $\sum_{j=1}^n \lambda_j^2 \leq (\operatorname{Tr}(M))^2/(n-1)$ implying 2-incoherence of M . When combined with Equation (62), this tells us that if

$$\|M - I\|_F^2 \leq n - 2 \operatorname{Tr}(M) + \frac{(\operatorname{Tr}(M))^2}{d-1} \quad (63)$$

then M is 2-incoherent (and thus k -incoherent for all $k \in \{2, \dots, n\}$).

If we treat the right-hand-side of Inequality (63) as a function of $\operatorname{Tr}(M)$ then standard calculus techniques show that it is minimized when $\operatorname{Tr}(M) = n-1$, and its minimum value is

$$n - 2(n-1) + \frac{(n-1)^2}{n-1} = 1. \quad (64)$$

It follows that if $\|M - I\|_F^2 \leq 1$ then Inequality (63) must hold, so M must be 2-incoherent. Taking the square root of both sides of this inequality gives the statement of the theorem. \square

Corollary 32. *For any unit vector $u \in \mathbb{C}^n$ supported on at most k coordinates, and any $x \geq 0$, it holds that*

$$D_x := (1 - \frac{\delta}{n})(x \frac{I}{n} + (1-x)uu^*) \in S(\mathcal{I}_k^n, -\delta) \quad (65)$$

for any $\delta \leq \frac{x}{2n}$.

Proof. Let $D_x + \delta E \in S(D, \delta)$, where E is an arbitrary Hermitian matrix with $\|E\|_F = 1$. Note that $\operatorname{Tr}(E) \leq n$, so the trace of $D_x + \delta E$ is at most 1. We need to prove that $D_x + \delta E \in \mathcal{I}_k^n$. Note that $D_x + \delta E = x((1 - \frac{\delta}{n})\frac{I}{n} + \frac{\delta}{x}E) + (1 - \frac{\delta}{n})(1-x)uu^*$. If $\delta \leq \frac{1}{2}n^{-1}x$, then $\delta(x^{-1} + 1) \leq n^{-1}$, so by Lemma 31 the matrix $((1 - \frac{\delta}{n})\frac{I}{n} + \frac{\delta}{x}E)$ is positive semidefinite of factor width $\leq k$. Since uu^* also has factor width $\leq k$, and \mathcal{I}_k^n is convex, the statement follows. \square

Proof of Lemma 22. The above corollary demonstrates that $B(I/n, n^{-1}) \subseteq \mathcal{I}_k^n \subseteq B(0, 1)$. Since the ratio of the radii of these two balls is polynomial, by Fact 20, $\operatorname{WeakOPT}(\mathcal{I}_k^n, 1/\text{poly})$ reduces to $\operatorname{WeakMEM}(\mathcal{I}_k^n, 1/\text{poly})$. This completes the proof. \square

D Proof of Lemma 23

Lemma 23 states that the set of extreme points of \mathcal{I}_k^n is

$$P = \{vv^* : v \text{ has at most } k \text{ non-zero entries and } \|v\|_2 = 1\} \cup \{0\}. \quad (66)$$

For the inclusion \supseteq , it is clear that 0 is an extreme point. Now, let vv^* be another point in P , and suppose that $vv^* = p\rho + (1-p)\sigma$ for some $\rho, \sigma \in \mathcal{I}_k^n$. Then $\rho = \alpha vv^*$ and $\sigma = \beta vv^*$ for

some $\alpha, \beta \in [0, 1]$ (otherwise vv^* would have rank > 1 , a contradiction). Taking the trace, we have $p\alpha + (1-p)\beta = 1$, so $\alpha = \beta = 1$.

For the inclusion \subseteq , suppose that ρ is an extreme point of \mathcal{I}_k^n . Let

$$\rho = \sum_{i=1}^m v_i v_i^* = \sum_{i=1}^m \frac{\text{Tr}(\rho)}{\|v_i\|^2} (v_i v_i^*) \quad (67)$$

be a factor width k decomposition of ρ , as in (8). Since this is a convex combination of elements of \mathcal{I}_k^n , it must hold that $m = 1$, i.e. $\rho = v_1 v_1^*$. It remains to prove that $\text{Tr}(\rho) \in \{0, 1\}$. Suppose toward contradiction that $\text{Tr}(\rho) = \alpha \in (0, 1)$. Then $\rho = \frac{1}{2}(\frac{1}{\alpha}\rho) + \frac{1}{2}((2\alpha - 1)\rho)$ is an expression of ρ as a non-trivial convex combination of elements of \mathcal{I}_k^n , hence ρ is not extremal. This completes the proof.

E Proof of Proposition 26

Here we prove Proposition 26. For convenience, we recall the statement of the proposition.

Proposition 33 (Proposition 26). *Let $n \geq k \geq 2$ be integers, let C be the adjacency matrix of a simple nonempty undirected graph G on n vertices with e edges, and let $C' = \frac{1}{\sqrt{2e}}C$ be the normalization in Frobenius norm. Then there exist functions $\gamma = \gamma(e, k, n) \in \mathbb{R}$ and $\delta = 1/\text{poly}(n)$ for which*

- $\mu(k, C') \geq \gamma + \delta$ if G contains a k -clique.
- $\mu(k, C') \leq \gamma - \delta$ if G does not contain a k -clique.

Proof. By Corollary 24, $\mu(k, C)$ is equal to the maximum eigenvalue of the adjacency matrix of an at-most- k -vertex subgraph of G . It is a standard fact that the maximum eigenvalue of the adjacency matrix of a k -vertex graph H with f edges is upper bounded by $\sqrt{\frac{2f(k-1)}{k}}$, with equality if and only if H is the complete graph (see e.g. [50]). Note that this quantity is equal to $k-1$ if H is the complete graph, and is upper-bounded by

$$\sqrt{(k-1)^2 - 2 + \frac{2}{k}} \leq k-1 - \frac{1}{\epsilon(k)} \quad (68)$$

otherwise, for some $\epsilon(k) = 1/\text{poly}(k)$.

Suppose that G contains a k -clique, and let $u \in \mathbb{C}^n$ be a unit vector supported on at most k coordinates for which it holds that $\text{Tr}(C'uu^*) = \frac{k-1}{\sqrt{2e}}$ (such a vector exists by extremality, see Lemma 23). Keeping δ indeterminate for now, let $x = 2n\delta$ and let $D_x \in S(\mathcal{I}_k^n, -\delta)$ as in (65). Note that

$$\text{Tr}(C'D_x) = (1 - \frac{\delta}{n}) \frac{(1-x)(k-1)}{e^2} \geq \frac{(1 - (2n+1)\delta)(k-1)}{\sqrt{2e}}. \quad (69)$$

Now suppose that G does not contain a k -clique. Let $X \in S(\mathcal{I}_k^n, \delta)$, so $X = F + \delta E$ for some $F \in \mathcal{I}_k^n$ and Hermitian matrix E of Frobenius norm 1. Then

$$\text{Tr}(C'X) = \text{Tr}(C'F) + \delta \text{Tr}(C'E) \leq \frac{k-1}{\sqrt{2e}} - \frac{1}{\epsilon\sqrt{2e}} + \delta. \quad (70)$$

Let $\gamma = \frac{k-1}{\sqrt{2e}} - \frac{1}{\epsilon\sqrt{2e}} + 2\delta$. Then $\text{Tr}(C'X) \leq \gamma - \delta$. Now, we just need to show that we can choose $\delta = 1/\text{poly}(n)$ small enough so that $\text{Tr}(C'D_x) \geq \gamma + \delta$. One can easily verify that the choice

$$\delta = \frac{1}{\epsilon\sqrt{2e}(3 + \frac{(2n+1)(k-1)}{\sqrt{2e}})} \geq \frac{1}{\text{poly}(n)} \quad (71)$$

suffices. This completes the proof. □