Distributional property testing in a quantum world

András Gilyén* Tongyang Li†

February 5, 2019

Abstract

A fundamental problem in statistics and learning theory is to test properties of distributions. We show that quantum computers can solve such problems with significant speed-ups. In particular, we give fast quantum algorithms for testing closeness between unknown distributions, testing independence between two distributions, and estimating the Shannon / von Neumann entropy of distributions. The distributions can be either classical or quantum, however our quantum algorithms require coherent quantum access to a process preparing the samples. Our results build on the recent technique of quantum singular value transformation, combined with more standard tricks such as divide-and-conquer. The presented approach is a natural fit for distributional property testing both in the classical and the quantum case, demonstrating the first speed-ups for testing properties of density operators that can be accessed coherently rather than only via sampling; for classical distributions our algorithms significantly improve the precision dependence of some earlier results.

1 Introduction

Distributional property testing is a fundamental problem in theoretical computer science (see, e.g. Goldreich (2017)). In such property testing questions the goal is to determine properties of probability distributions with the least number of independent samples. This has intimate connections and applications to statistics, learning theory, and algorithm design.

The merit of distributional property testing mainly comes from the fact that the testing of many properties admits *sublinear* algorithms. For instance, given the ability to take samples from a discrete distribution p on $[n] := \{1, \dots, n\}$, it requires $\Theta(n/\epsilon^2)$ samples to "learn" p, i.e., to construct a distribution q on $[n]$ such that $\|p - q\|_1 \leq \epsilon$ with success probability at least $2/3$ ($\|\cdot\|_1$ being ℓ^1-distance). However, testing whether $p = q$ or $\|p - q\|_1 > \epsilon$ requires only $\Theta(\max\{\frac{n^{2/3}}{\epsilon^{4/3}}, \frac{n^{1/2}}{\epsilon^2}\})$ samples from p and q (Chan et al. (2014)), which is sublinear in n and significantly smaller than the complexity of learning the entire distributions. See Section 1.4 for more examples and discussions.

In this paper, we study the impact of quantum computation on distributional property testing problems. We are motivated by the emerging topic of "quantum property testing" (see the survey of Montanaro and de Wolf (2016)) which focuses on investigating the quantum advantage in testing classical statistical properties. Quantum speed-ups have already been established for a

*QuSoft, CWI and University of Amsterdam, the Netherlands. Supported by ERC Consolidator Grant QPROGRESS and partially supported by QuantERA project QuantAlgo 680-91-034. gilyen@cwi.nl

†Department of Computer Science, Institute for Advanced Computer Studies, and Joint Center for Quantum Information and Computer Science, University of Maryland. Supported by IBM PhD Fellowship, QISE-NET Triplet Award (NSF DMR-1747426), and the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Algorithms Teams program. tongyang@cs.umd.edu

few specific problems such as testing closeness between distributions (Bravyi et al. (2011); Montanaro (2015)), testing identity to known distributions (Chakraborty et al. (2010)), estimating entropies (Li and Wu (2018)), etc. In this paper we propose a generic approach for quantum distributional property testing, and illustrate its power on a few examples. This is our attempt to make progress on the question:

Can quantum computers test properties of distributions systematically and more efficiently?

1.1 Problem statements

Throughout the paper, we denote probability distributions on $[n]$ by p and q; their ℓ^α-distance is defined as $\|p-q\|_\alpha := (\sum_{i=1}^n |p_i - q_i|^\alpha)^{\frac{1}{\alpha}}$. Similarly, we denote $n \times n$ density operators[1] (=quantum distributions) by ρ and σ; their ℓ^α-distance is defined via the corresponding Schatten norm.

Input models. To formulate the problems we address, we define classical and quantum access models for distributions on $[n]$. We begin with the very natural model of sampling.

Definition 1 (Sampling)**.** *A classical distribution $(p_i)_{i=1}^n$ is accessible via* classical sampling *if we can request samples from the distribution, i.e., get a random $i \in [n]$ with probability p_i. A quantum distribution $\rho \in \mathbb{C}^{n\times n}$ is accessible via* quantum sampling *if we can request copies of the state ρ.*

Now we define a coherent analogue of the above sampling model. To our knowledge this type of query-access was not studied before in detail, especially in the context of density operator testing. The motivation for this input model is the following: we can think about a density operator as the outcome of some physical process. If we are able to simulate the corresponding process on a fault-tolerant quantum computer, then it provides purified access to the density operator. In the special case when we study a classical probability distribution coming from some classical randomized process, we can simply simulate the classical randomized process on a quantum computer.

Definition 2 (Purified quantum query-access)**.** *A density operator $\rho \in \mathbb{C}^{n\times n}$, has* purified quantum query-access *if we have access to a unitary oracle U_ρ (and its inverse) acting as*[2]

$$U_\rho|0\rangle_A|0\rangle_B = |\psi_\rho\rangle_{AB} = \sum_{i=1}^n \sqrt{p_i}|\phi_i\rangle_A|\psi_i\rangle_B, \text{ (where } \langle\phi_i|\phi_j\rangle = \langle\psi_i|\psi_j\rangle = \text{(Kronecker) } \delta_{ij})$$

such that $\mathrm{Tr}_A(|\psi_\rho\rangle\langle\psi_\rho|) = \rho$. If $|\psi_i\rangle = |i\rangle$, then $\rho = \sum_i^n p_i|i\rangle\langle i|$ is a diagonal density operator which can be identified with the classical distribution p, so we can simply write U_p instead of U_ρ. With a slight abuse of notation sometimes we will concisely write $|\rho\rangle$ instead of $|\psi_\rho\rangle$.

We also define an even stronger input model that is considered in a series of earlier works, see, e.g., (Bravyi et al. (2011); Chakraborty et al. (2010); Li and Wu (2018); Bun et al. (2018)).

[1] For readers less familiar with quantum computing, a density operator (=quantum distribution) $\rho \in \mathbb{C}^{n\times n}$ is a positive semidefinite matrix with $\mathrm{Tr}[\rho] = 1$. Please refer to the textbook Nielsen and Chuang (2000) for more information.

[2] $|\psi\rangle \in \mathbb{C}^n$ denotes a "ket" vector and $\langle\psi| = (|\psi\rangle)^\dagger$ stands for its conjugate transpose, called "bra" in Dirac notation; $|i\rangle = \vec{e}_i$ is the i^{th} basis vector. An ℓ^2-normalized $|\psi\rangle$ is called a pure state, and corresponds to density operator $|\psi\rangle\langle\psi|$. For $A = \mathbb{C}^k, B = \mathbb{C}^n$ and $|\phi\rangle \in A \otimes B$ we denote by $\mathrm{Tr}[|\phi\rangle\langle\phi|]_A \in B \otimes B^* = \mathbb{C}^{n\times n}$ the partial trace over A.

Definition 3 (Classical distribution with discrete query-access)**.** *A classical distribution* $(p_i)_{i=1}^n$*, has* discrete query-access *if we have classical / quantum query-access to a function* $f\colon S \to [n]$ *such that for all* $i \in [n]$*,* $p_i = |\{s \in [S] : f(s) = i\}|/S$*. (Typically the interesting regime is when* $|S| \gg n$*.) In the quantum case a query oracle is a unitary operator* O *acting on* $\mathbb{C}^{|S|} \otimes \mathbb{C}^n$ *as*

$$\mathrm{O}\colon |s, 0\rangle \leftrightarrow |s, f(s)\rangle \text{ for all } s \in S.$$

Note that if one first creates a uniform superposition over S and then makes a query, then the above oracle turns into a purified query oracle to a classical distribution as in Definition 2. Therefore all lower bounds that are proven in this model also apply to the purified query-access oracles. In fact all algorithms that the authors are aware of do this conversion, so they effectively work in the purified query-access model. Moreover, we conjecture that the two input models are equivalent when $|S| \gg n$. For this reason we only work with the purified query-access model in this work.

Another strengthening of the purified query-access model for classical distributions when we have access to a unitary (and its inverse) acting as $|0\rangle \mapsto \sum_{i=1}^n \sqrt{p_i}|i\rangle$.

Definition 4 (Classical distribution with pure-state preparation access)**.** *A classical distribution* $(p_i)_{i=1}^n$*, is accessible via* pure state preparation oracle *if we have access to a unitary oracle* U_{pure} *(and its inverse) acting as*

$$U_{pure}\colon |0\rangle \mapsto \sum_{i=1}^{n} \sqrt{p_i}|i\rangle.$$

This is again strictly stronger than the purified query-access model. In order to simulate purified queries we can first do a pure state query and then copy $|i\rangle$ to a second fresh ancillary register using, e.g., some CNOT gates. Finally, for completeness we mention that one could also consider a model similar to the above where one can only request samples of pure states of the form $\sum_{i=1}^n \sqrt{p_i}|i\rangle$, as studied for example in Arunachalam and de Wolf (2017); Arunachalam et al. (2018).

We will mostly focus on the first two input models and will only use the latter strengthenings of the purified query-access model for invoking and proving lower bounds.

Property testing problems. We study three distributional properties: ℓ^α-closeness testing, independence testing, and entropy estimation. These properties are highly-representative; classically, these testers motivate general algorithms for testing properties of discrete distributions (Diakonikolas and Kane (2016); Acharya et al. (2017a)).

For brevity we only give the definitions for classical distributions; similar definitions apply to quantum density matrices if we replace vector norms by the corresponding Schatten norms.

Definition 5 (ℓ^α-closeness testing)**.** *Given* $\epsilon > 0$ *and two probability distributions* p*,* q *on* $[n]$*,* ℓ^α-closeness testing *is to decide whether* $p{=}q$ *or* $\|p{-}q\|_\alpha \geq \epsilon$ *with success probability at least* $\frac{2}{3}$*. Robust testing: decide whether* $\|p{-}q\|_\alpha {\leq} 0.99\epsilon$ *or* $\|p{-}q\|_\alpha \geq \epsilon$ *with success probability at least* $\frac{2}{3}$*.*

Definition 6 (Independence testing)**.** *Given* $\epsilon > 0$ *and a probability distribution* p *on* $[n] \times [m]$ *with* $n \geq m$*,* independence testing *is to decide, with success probability at least* $\frac{2}{3}$*, whether* p *is a product distribution or* p *is* ϵ*-far in* ℓ^1*-norm from any product distribution on* $[n] \times [m]$ *.*

Definition 7 (Entropy estimation)**.** *Given* $\epsilon > 0$ *and a density operator* $\rho \in \mathbb{C}^{n\times n}$*,* entropy estimation *is to estimate the Shannon / von Neumann entropy* $H(\rho) = -\mathrm{Tr}[\rho \log(\rho)]$ *within additive* ϵ*-precision with success probability at least* $\frac{2}{3}$*.*

1.2 Contributions

We give a systematic study of distributional property testing for classical / quantum distributions, and obtain the following results for the purified quantum query model of Definition 2:

- Entropy estimation of classical / quantum distributions costs $\widetilde{\mathcal{O}}\left(\frac{\sqrt{n}}{\epsilon^{1.5}}\right)$ and $\widetilde{\mathcal{O}}\left(\frac{n}{\epsilon^{1.5}}\right)$ queries respectively, as we prove in Theorem 12 and Theorem 13.
- Robust ℓ^2-closeness testing of classical / quantum distributions costs $\tilde{\Theta}\left(\frac{1}{\epsilon}\right)$ and $\mathcal{O}\left(\min\left(\frac{\sqrt{n}}{\epsilon}, \frac{1}{\epsilon^2}\right)\right)$ queries respectively, as we prove in Theorem 14 and Theorem 15.
- ℓ^1-closeness testing of classical / quantum distributions costs $\widetilde{\mathcal{O}}\left(\frac{\sqrt{n}}{\epsilon}\right)$ and $\mathcal{O}\left(\frac{n}{\epsilon}\right)$ queries respectively, as we prove in Corollary 17.
- Independence testing of classical / quantum distributions costs $\widetilde{\mathcal{O}}\left(\frac{\sqrt{nm}}{\epsilon}\right)$ and $\mathcal{O}\left(\frac{nm}{\epsilon}\right)$ queries respectively, as we prove in Corollary 18.

For context, we compare our results with previous classical and quantum results in Table 1 and Table 2. (Note that all of our results are gate efficient, because they are based on singular value transformation and amplitude estimation, both of which have gate-efficient implementations.)

problem / model	ℓ^1-closeness testing	(robust) ℓ^2-closeness testing	Shannon / von Neumann entropy
Classical sampling	$\Theta\left(\max\left\{\frac{n^{2/3}}{\epsilon^{4/3}}, \frac{n^{1/2}}{\epsilon^2}\right\}\right)$ Chan et al. (2014)	$\Theta\left(\frac{1}{\epsilon^2}\right)$ Chan et al. (2014)	$\Theta\left(\frac{n}{\epsilon \log n} + \frac{\log^2 n}{\epsilon^2}\right)$ Jiao et al. (2015), Wu and Yang (2016)
Classical with quantum query-access	$\boldsymbol{\widetilde{\mathcal{O}}\left(\frac{\sqrt{n}}{\epsilon}\right)}$	$\boldsymbol{\tilde{\Theta}\left(\frac{1}{\epsilon}\right)}$	$\boldsymbol{\widetilde{\mathcal{O}}\left(\frac{\sqrt{n}}{\epsilon^{1.5}}\right)}$; $\tilde{\Omega}(\sqrt{n})$ Bun et al. (2018)
Quantum state with purification	$\boldsymbol{\mathcal{O}\left(\frac{n}{\epsilon}\right)}$	$\boldsymbol{\mathcal{O}\left(\min\left(\frac{\sqrt{n}}{\epsilon}, \frac{1}{\epsilon^2}\right)\right)}$	$\boldsymbol{\widetilde{\mathcal{O}}\left(\frac{n}{\epsilon^{1.5}}\right)}$
Quantum state sampling	$\Theta\left(\frac{n}{\epsilon^2}\right)$ Bădescu et al. (2017)	$\Theta\left(\frac{1}{\epsilon^2}\right)$ Bădescu et al. (2017)	$\mathcal{O}\left(\frac{n^2}{\epsilon^2}\right)$, $\Omega\left(\frac{n^2}{\epsilon}\right)$ Acharya et al. (2017b)

Table 1: Summary of sample and query complexity results. Our new bounds are printed in **bold**. For classical distributions with quantum query-access[3] we prove (almost) matching upper and lower bounds for ℓ^2-testing, and improve the previous best complexity $\widetilde{\mathcal{O}}\left(\sqrt{n}/\epsilon^{2.5}\right)$ for ℓ^1-testing by Montanaro (2015) and $\widetilde{\mathcal{O}}\left(\sqrt{n}/\epsilon^2\right)$ for Shannon entropy estimation by Li and Wu (2018). We are not aware of prior work on testing quantum distributions with purified query-access.

	Sample complexity	(Purified) Query complexity
Classical	$\Theta\left(\frac{n}{\log n}\right)$ Valiant and Valiant (2011a)	$\widetilde{\Theta}(\sqrt{n})$ Li and Wu (2018); Bun et al. (2018)
Quantum	$\Theta\left(n^2\right)$ Acharya et al. (2017b)	$\boldsymbol{\widetilde{\mathcal{O}}(n)}$

Table 2: Complexities of Shannon / von Neumann entropy estimation with constant precision. It seems that the n-dependence is roughly quadratically higher for quantum distributions, while coherent quantum access gives a quadratic advantage for both classical and quantum distributions. This suggests that our entropy estimation algorithm has essentially optimal n-dependence for density operators with purified access, however we do not have a matching lower bound yet.

[3] Recent results of Chailloux (2018) imply that in this model quantum speed-ups are at most cubic.

1.3 Techniques

The motivating idea behind our approach is that if we can prepare a purification of a quantum distribution / density operator ρ, then we can construct a unitary U, which has this density operator in the top-left corner, using only two queries to U_ρ. This observation is originally due to Low and Chuang (2016). We call such a unitary a *block-encoding* of ρ:

$$U = \begin{bmatrix} \rho & \cdot \\ \cdot & \cdot \end{bmatrix} \iff \rho = \big(\langle 0|^{\otimes a} \otimes I\big) U \big(|0\rangle^{\otimes a} \otimes I\big).$$

One can think of a block-encoding as a probabilistic implementation of the linear map ρ: given an input state $|\psi\rangle$, applying the unitary U to the state $|0\rangle^{\otimes a}|\psi\rangle$, measuring the first a-qubit register and post-selecting on the $|0\rangle^{\otimes a}$ outcome, we get a state $\propto \rho|\psi\rangle$ in the second register. Block-encodings are easy to work with, for example given a block-encoding of ρ and σ we can easily construct a block-encoding of $(\rho-\sigma)/2$, see for example in the work of Chakraborty et al. (2018).

Example application to ℓ^3-testing. The problem is to decide whether $\rho = \sigma$ or $\|\rho - \sigma\|_3 \geq \epsilon$, with query complexity $\mathcal{O}\Big(\epsilon^{-\frac{3}{2}}\Big)$. The first idea is that if we can prepare a purification of ρ and σ, then by flipping a fair coin and preparing ρ or σ based on the outcome, we can also prepare a purification of $(\rho+\sigma)/2$. The second idea is to combine the block-encodings of ρ and σ to apply the map $\frac{\rho-\sigma}{2}$ to the purification of $(\rho+\sigma)/2$, to get

$$\left|\frac{\rho+\sigma}{2}\right\rangle \mapsto \left(\frac{\rho-\sigma}{2} \otimes I\right)\left|\frac{\rho+\sigma}{2}\right\rangle|0\rangle + \ldots|1\rangle.$$

Finally, apply amplitude estimation with setting $M = \Theta(\epsilon^{-\frac{3}{2}})$. This works since if $\|\rho-\sigma\|_3 \geq \epsilon$, then the $|0\rangle$ ancilla state has probability $\mathrm{Tr}\big[(\rho-\sigma)^2(\rho+\sigma)\big]/8 \geq \mathrm{Tr}\big[|\rho-\sigma|^3\big]/8 \geq \epsilon^3/8$.

Working with singular values. The above is a promising approach because it directly makes the density operator in question operationally accessible. However, it turns out that using this simple block-encodings is often suboptimal for distribution testing, because a query in some sense gives access to the square-root of ρ, whereas this unitary has ρ itself in the top-left corner. Since the problems often heavily depend on smaller eigenvalues of ρ, the square root of ρ is more desirable since it has quadratically larger singular-/eigenvalues.

Therefore, we show how to efficiently construct a unitary matrix whose top-left corner contains a matrix with singular values $\sqrt{p_1}, \ldots, \sqrt{p_n}$, given purified access to a classical distribution p. To be more precise, we define *projected unitary encodings*, which represents a matrix A in the form of $\Pi U \widetilde{\Pi}$, where $\Pi, \widetilde{\Pi}$ are orthogonal projectors and U is a unitary matrix. One can think about U in a projected unitary encoding as a probabilistic implementation of the map $A\colon \mathrm{img}\Big(\widetilde{\Pi}\Big) \to \mathrm{img}(\Pi)$. Take for example $U := (U_p \otimes I)$, $\Pi := (\sum_{i=1}^n I \otimes |i\rangle\langle i| \otimes |i\rangle\langle i|)$, and $\widetilde{\Pi} := (|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes I)$. As we show in Appendix A these operators form a projected unitary encoding of

$$A = \Pi U \widetilde{\Pi} = \sum_{i=1}^{n} \sqrt{p_i}|\phi_i\rangle\langle 0| \otimes |i\rangle\langle 0| \otimes |i\rangle\langle i|. \tag{1}$$

We can use a similar trick for a general density operator ρ too. However, there is a major difficulty which arises from the fact that we do not a prioiri know the diagonalizig basis of ρ. Therefore we use slightly different operators. Let W be a unitary,[4] mapping $|0\rangle|0\rangle \mapsto$

[4]This unitary is easy to implement, e.g., by using a few Hadamard and CNOT gates.

$\sum_{j=1}^{n} \frac{|j\rangle|j\rangle}{\sqrt{n}}$. Let $U' := \Big(I \otimes U_\rho^\dagger\Big)\big(W^\dagger \otimes I\big)$, $\Pi' := (I \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|)$ and $\widetilde{\Pi}$ as above. As we show in Appendix A these operators form a projected unitary encoding of

$$A' = \Pi U' \widetilde{\Pi} = \sum_{i=1}^{n} \sqrt{\frac{p_i}{n}} |\phi_i'\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes |0\rangle\langle \psi_i|. \tag{2}$$

As we can see, the case of general density operators is less efficient, it only gives operational access to the "square root" of ρ/n. If the $1/\sqrt{n}$ factor could be directly improved, that would speed up our von Neumann entropy estimation algorithm Theorem 13, which seems unlikely, cf. Table 2.

General recipe. Our recipe to distributional property testing can be summarized as follows.

1.) Construct a unitary matrix / quantum circuit operationally representing the distribution.

2.) Transform the singular values of the corresponding matrix according to a desired function.

3.) Apply the resulting map to the purification of the distribution, or another suitable state.

4.) Estimate the amplitude of the flagged output state and conclude.

The above general scheme describes our approach to the problems we discuss in this paper. Sometimes it is useful to divide the probabilities / singular values into bins, and fine-tune the algorithm by using the approximate knowledge of the size of the singular values. This divide-and-conquer strategy is at the core of our improved robust ℓ^2-closeness tester of Theorem 14.

1.4 Related works on distributional property testing

Classical algorithms. Many distributional property testing problems fall into the category of *closeness testing*, where we are given the ability to take independent samples from two unknown distributions p and q with cardinality n, and the goal is to determine whether they are the same versus significantly different. For ℓ^1-*closeness testing*, which is about testing whether $p = q$ or $\|p - q\|_1 \geq \epsilon$, Batu et al. (2013) first gave a sublinear algorithm using $\tilde{O}(n^{2/3}/\epsilon^{8/3})$ samples to p and q. The follow-up work by Chan et al. (2014) determined the optimal sample complexity as $\Theta(\max\{\frac{n^{2/3}}{\epsilon^{4/3}}, \frac{n^{1/2}}{\epsilon^2}\})$; the same paper also gave a tight bound $\Theta(\frac{1}{\epsilon^2})$ for ℓ^2-*closeness testing*.

Besides closeness testing, a similar problem is *identity* testing where one of the distributions, say q, is known and we are given independent samples from the other distribution p. For ℓ^1 *identity testing*, it is known that the sample complexity can be smaller than that of ℓ^1-closeness testing, which was proved by Batu et al. (2001) to be $\tilde{O}(\sqrt{n}/\epsilon^4)$ and then Paninski (2008) gave the tight bound $\Theta(\sqrt{n}/\epsilon^2)$. More recently, Diakonikolas and Kane (2016) proposed a modular reduction-based approach for distributional property testing problems, which recovered all closeness and identity testing results above. Furthermore, they also studied *independence testing*, i.e., whether a distribution on $[n]\times[m]$ $(n \geq m)$ is a product distribution or at least ϵ-far in ℓ^1-distance from any product distribution, and determined the optimal bound $\Theta(\max\{\frac{n^{2/3}m^{1/3}}{\epsilon^{4/3}}, \frac{(nm)^{1/2}}{\epsilon^2}\})$.

Apart from the relationship between distributions, properties of a single distribution also have been extensively studied. One of the most important properties is *Shannon entropy* (Shannon (1948)) because it measures for example compressibility. The sample complexity of estimating $H(p)$ within additive error ϵ has been intensively studied (Batu et al. (2005); Paninski (2003, 2004)); in particular, Valiant and Valiant (2011a,b) gave an explicit algorithm for entropy estimation using $\Theta(\frac{n}{\epsilon \log n})$ samples when $\epsilon = \Omega(n^{-0.03})$ and $\epsilon = O(1)$; for the general case Jiao et al. (2015) and Wu and Yang (2016) gave the optimal estimator with $\Theta\Big(\frac{n}{\epsilon \log n} + \frac{(\log n)^2}{\epsilon^2}\Big)$ samples.

Quantum algorithms. The first paper on distributional property testing by quantum algorithms was by Bravyi et al. (2011), which considered classical distributions with discrete quantum query-access (see Definition 3); it gives a quantum query complexity upper bound $O(\sqrt{n}/\epsilon^6)$ for ℓ^1-closeness testing and $O(n^{1/3}/\epsilon^{4/3})$ for identity testing to the uniform distribution on $[n]$. Subsequently, Chakraborty et al. (2010) gave an algorithm for identity testing (to an arbitrary known distribution) with $\widetilde{\mathcal{O}}\big(n^{1/3}/\epsilon^5\big)$ queries, and Montanaro (2015) improved the ϵ-dependence of ℓ^1-closeness testing to $\widetilde{\mathcal{O}}\big(\sqrt{n}/\epsilon^{2.5}\big)$. More recently, Li and Wu (2018) studied entropy estimation under this model, and gave a quantum algorithm for Shannon entropy estimation with $\widetilde{\mathcal{O}}\big(\sqrt{n}/\epsilon^2\big)$ queries and also sublinear quantum algorithms for estimating Rényi entropies (Rényi (1961)).

Another type of quantum property testing results (O'Donnell and Wright (2015, 2016, 2017); Bădescu et al. (2017); Acharya et al. (2017b)) concern *density matrices*, where the ℓ^1-distance becomes the trace distance and the Shannon entropy becomes the von Neumann entropy. To be more specific, for n-dimensional density matrices, the number of samples needed for ℓ^1 and ℓ^2-closeness testing are $\Theta(n/\epsilon^2)$ and $\Theta(1/\epsilon^2)$ (Bădescu et al. (2017)), respectively. In addition Acharya et al. (2017b) gave upper and lower bounds $\mathcal{O}\big(n^2/\epsilon^2\big), \Omega\big(n^2/\epsilon\big)$ for estimating the von Neumann entropy of an n-dimensional density matrix with accuracy ϵ.

1.5 Organization of the paper

The rest of the paper is organized as follows. In Section 2 we introduce two important quantum algorithmic techniques, amplitude estimation and singular value transformation. We give entropy estimators of classical and quantum distributions in Section 3. In Section 4 we give an (essentially) optimal quantum algorithm for robustly testing ℓ^2-closeness of classical distributions, and another efficient robust ℓ^2-closeness tester for quantum distributions. Proof details of projected encodings, polynomial approximations for singular value transformation, and corollaries about ℓ^1-closeness and independence testing are deferred to Appendix A, B, and C respectively.

2 Preliminaries

2.1 Amplitude estimation

Classically, given i.i.d. samples of a Bernoulli random variable X with $\mathbb{E}[X] = p$, it takes $\Theta(1/\epsilon^2)$ samples to estimate p within ϵ with high success probability. Quantumly, if we are given a unitary U such that

$$U|0\rangle|0\rangle = \sqrt{p}|0\rangle|\phi\rangle + |0^\perp\rangle, \quad \text{where } \||\phi\rangle\| = 1 \text{ and } (\langle 0| \otimes I)|0^\perp\rangle = 0, \tag{3}$$

then if measure the output state, we get 0 in the first register with probability p. Given access to U we can estimate the value of p quadratically more efficiently than what is possible by sampling:

Theorem 8. (Brassard et al., 2002, Theorem 12) *Given U satisfying (3), the amplitude estimation algorithm outputs $\tilde{p}$ such that $\tilde{p} \in [0,1]$ and*

$$|\tilde{p} - p| \leq \frac{2\pi\sqrt{p(1-p)}}{M} + \frac{\pi^2}{M^2} \tag{4}$$

with success probability at least $8/\pi^2$, using M calls to U and $U^\dagger$.

In particular, if we take $M = \left\lceil 2\pi\left(\frac{2\sqrt{p}}{\epsilon} + \frac{1}{\sqrt{\epsilon}}\right)\right\rceil = \Theta\left(\frac{\sqrt{p}}{\epsilon} + \frac{1}{\sqrt{\epsilon}}\right)$ in (4), we have

$$|\tilde{p} - p| \leq \frac{2\pi\sqrt{p(1-p)}}{2\pi}\epsilon + \frac{\pi^2}{4\pi^2}\epsilon^2 \leq \frac{\epsilon}{2} + \frac{\epsilon}{4} \leq \epsilon.$$

Therefore, using only $\Theta(1/\epsilon)$ implementations of U and $U^\dagger$, we could get an ϵ-additive approximation of p with success probability at least $8/\pi^2$, which is a quadratic speed-up compared to the classical sample complexity $\Theta(1/\epsilon^2)$. The success probability can be boosted to $1-\nu$ by executing the algorithm for $\Theta(\log 1/\nu)$ times and taking the median of the estimates.

2.2 Quantum singular value transformation

Singular value decomposition (SVD) is one of the most important tools in linear algebra, generalizing eigen-decomposition of Hermitian matrices. Recently, Gilyén et al. (2018) proposed *quantum singular value transformation* which turns our to be very useful for property testing. Mathematically, it is defined as follows:

Definition 9 (Singular value transformation)**.** *Let $f : \mathbb{R} \to \mathbb{C}$ be an even or odd function. Let $A \in \mathbb{C}^{\tilde{d}\times d}$ have the following singular value decomposition*

$$A = \sum_{i=1}^{d_{\min}} \varsigma_i |\tilde{\psi}_i\rangle\langle\psi_i|,$$

where $d_{\min} := \min(d, \tilde{d})$. For the function f we define the singular value transformation *on A as*

$$f^{(SV)}(A) := \begin{cases} \sum_{i=1}^{d_{\min}} f(\varsigma_i)|\tilde{\psi}_i\rangle\langle\psi_i| & \text{if } f \text{ is odd, and} \\ \sum_{i=1}^{d} f(\varsigma_i)|\psi_i\rangle\langle\psi_i| & \text{if } f \text{ is even, where for } i \in [d]\setminus[d_{\min}] \text{ we define } \varsigma_i := 0. \end{cases}$$

Quantum singular value transformation by real polynomials can be efficiently implemented on a quantum computer as follows:

Theorem 10. (Gilyén et al., 2018, Corollary 18) *Let $\mathcal{H}_U$ be a finite-dimensional Hilbert space and let $U, \Pi, \widetilde{\Pi} \in \mathrm{End}(\mathcal{H}_U)$ be linear operators on $\mathcal{H}_U$ such that U is a unitary, and $\Pi, \widetilde{\Pi}$ are orthogonal projectors. Suppose that $P = \sum_{k=0}^{n} a_k x^k \in \mathbb{R}[x]$ is a degree-n polynomial such that*

- *$a_k \neq 0$ only if $k \equiv n \mod 2$, and*
- *for all $x \in [-1, 1]$: $|P(x)| \leq 1$.*

Then there exist $\Phi \in \mathbb{R}^n$, such that

$$P^{(SV)}\left(\widetilde{\Pi}U\Pi\right) = \begin{cases} \left(\langle+|\otimes\widetilde{\Pi}\right)\left(|0\rangle\langle 0|\otimes U_\Phi + |1\rangle\langle 1|\otimes U_{-\Phi}\right)\left(|+\rangle\otimes\Pi\right) & \text{if } n \text{ is odd, and} \\ \left(\langle+|\otimes\Pi\right)\left(|0\rangle\langle 0|\otimes U_\Phi + |1\rangle\langle 1|\otimes U_{-\Phi}\right)\left(|+\rangle\otimes\Pi\right) & \text{if } n \text{ is even,} \end{cases}$$

where $U_\Phi = e^{i\phi_1(2\widetilde{\Pi}-I)}U\prod_{j=1}^{(n-1)/2}\left(e^{i\phi_{2j}(2\Pi-I)}U^\dagger e^{i\phi_{2j+1}(2\widetilde{\Pi}-I)}U\right)$.[5]

[5]This is the mathematical form for odd n; even n is defined similarly.

Thus for an even or odd polynomial P of degree n, we can apply singular value transformation of the matrix $\widetilde{\Pi}U\Pi$ with n uses of U, $U^\dagger$ and the same number of controlled reflections $I-2\Pi$, $I-2\widetilde{\Pi}$.

To apply singular value transformation corresponding to our problems, we need low-degree polynomial approximations to the following functions, which we construct in Appendix B.

Lemma 11. (Polynomial approximations) *Let $\beta \in (0,1]$, $\eta \in (0, \frac{1}{2}]$ and $t \geq 1$. There exists polynomials $\tilde{P}, \tilde{Q}, \tilde{S}$ such that*

- $\forall x \in [\frac{1}{t}, 1]$: $|\tilde{P}(x) - \frac{1}{2tx}| \leq \eta$, *and* $\forall x \in [-1,1]$: $-1 \leq \tilde{P}(x) = \tilde{P}(-x) \leq 1$,
- $\forall x \in [-\frac{1-\beta}{t}, \frac{1-\beta}{t}]$: $|\tilde{Q}(x) - tx| \leq \eta \cdot (tx)$, *and* $\forall x \in [-1,1]$: $\tilde{Q}(x) = -\tilde{Q}(-x) \leq 1$,
- $\forall x \in [\beta, 1]$: $|\tilde{S}(x) - \frac{\ln(1/x)}{2\ln(2/\beta)}| \leq \eta$, *and* $\forall x \in [-1,1]$: $-1 \leq \tilde{S}(x) = \tilde{S}(-x) \leq 1$,

moreover $\deg(\tilde{P}) = \mathcal{O}\Big(t \log\Big(\frac{1}{\eta}\Big)\Big)$, $\deg(\tilde{Q}) = \mathcal{O}\Big(\frac{t}{\beta} \log\Big(\frac{1}{\eta}\Big)\Big)$, *and* $\deg(\tilde{S}) = \mathcal{O}\Big(\frac{1}{\beta} \log\Big(\frac{1}{\eta}\Big)\Big)$.

3 Entropy estimation

3.1 Classical distributions with purified quantum query-access

Recall that we introduced purified quantum query-access in Definition 2. In particular, for a classical distribution p on $[n]$, we are given a unitary U_p acting on $\mathbb{C}^{n \times n}$ such that

$$U_p|0\rangle_A|0\rangle_B = |\psi_p\rangle = \sum_{i=1}^{n} \sqrt{p_i}|\phi_i\rangle_A|i\rangle_B. \tag{5}$$

We use U_p and $U_p^\dagger$ to estimate the Shannon entropy $H(p)$:

Theorem 12. *For any $0 < \epsilon < 1$, we can estimate $H(p)$ with accuracy ϵ with success probability at least $2/3$ using $\mathcal{O}\Big(\frac{\sqrt{n}}{\epsilon^{1.5}} \log^{1.5}\Big(\frac{n}{\epsilon}\Big) \log\Big(\frac{\log n}{\epsilon}\Big)\Big)$ calls to U_p and $U_p^\dagger$.*

Proof. The general idea is to first construct a unitary matrix with singular values $\sqrt{p_1}, \ldots, \sqrt{p_n}$. We use the construction of Eq. (1) and apply singular value transformation (Theorem 10) by a polynomial $\tilde{S}$ constructed in Corollary 11, setting $\eta = \frac{\epsilon}{24\ln(2/\beta)}$ and $\beta = \sqrt{\Delta}$ for $\Delta = \frac{\epsilon}{4n\ln(n/\epsilon)}$. Notice that this Δ satisfies

$$\Delta\Big(\ln\Big(\frac{1}{\Delta}\Big) + 1\Big) = \frac{\epsilon}{4n\ln(n/\epsilon)} \cdot \ln\frac{4en\ln(n/\epsilon)}{\epsilon} \leq \frac{\epsilon}{4n\ln(n/\epsilon)} \cdot \ln\frac{n^2}{\epsilon^2} = \frac{\epsilon}{2n}, \tag{6}$$

provided that $\frac{n}{\epsilon} \geq 42$. Note that the polynomial $\tilde{S}$ satisfies both conditions in Theorem 10. Applying the singular value transformed version of the operator (1) to the state $|\psi_p\rangle$ results in

$$|\widetilde{\Psi_p}\rangle = \sum_{i=1}^{n} \sqrt{p_i}\tilde{S}(\sqrt{p_i})|\phi_i\rangle_A|i\rangle_B|0\rangle + \ldots|1\rangle. \tag{7}$$

Preparing $|\widetilde{\Psi_p}\rangle$ costs $\deg\tilde{S} = \mathcal{O}\Big(\frac{1}{\beta}\log\Big(\frac{1}{\eta}\Big)\Big) = \mathcal{O}\Big(\sqrt{\frac{n}{\epsilon}\log\Big(\frac{n}{\epsilon}\Big)}\log\Big(\frac{\log n}{\epsilon}\Big)\Big)$ uses of U_p and $U_p^\dagger$ and the same number of controlled reflections through $\Pi, \widetilde{\Pi}$. Furthermore, Eq. (15) implies that for all i such that $p_i \geq \Delta$,

$$\Big|\frac{p_i\ln(1/p_i)}{4\ln(2/\beta)} - p_i\tilde{S}(\sqrt{p_i})\Big| = p_i \cdot \Big|\frac{\ln(1/\sqrt{p_i})}{2\ln(2/\beta)} - \tilde{S}(\sqrt{p_i})\Big| \leq \eta p_i. \tag{8}$$

For all i such that $p_i < \Delta$, we have

$$\left|\frac{p_i \ln(1/p_i)}{4\ln(2/\beta)} - p_i\tilde{S}(\sqrt{p_i})\right| \le \frac{p_i \ln(1/p_i) + p_i}{4\ln(2/\beta)} \le \frac{\Delta(\ln(\frac{1}{\Delta}) + 1)}{4\ln(2/\beta)} \le \frac{\epsilon}{8n\ln(2/\beta)}, \tag{9}$$

where the first inequality comes from the fact that $|\tilde{S}(x)| \le 1$ for all $x \in [-1, 1]$, the second inequality comes from the monotonicity of $x(\ln(1/x) + 1)$ on $(0, \frac{1}{\Delta}]$, and the third inequality comes from (6). As a result of (5), (8), and (9), we have

$$\begin{aligned}\left|(\langle\psi_p| \otimes \langle 0|)|\widetilde{\Psi_p}\rangle - \frac{H(p)}{4\ln(2/\beta)}\right| &= \left|p_i\tilde{S}(\sqrt{p_i}) - \sum_{i=1}^{n}\frac{p_i\log(1/p_i)}{4\ln(2/\beta)}\right| \\ &\le \sum_{i\colon p_i<\Delta}\frac{\epsilon}{8n\ln(2/\beta)} + \sum_{i\colon p_i\ge\Delta}\eta p_i \\ &\le \frac{\epsilon}{8\ln(2/\beta)} + \frac{\epsilon}{24\ln(2/\beta)} = \frac{\epsilon}{6\ln(2/\beta)}.\end{aligned}$$

Therefore, $|4\ln(2/\beta)(\langle\psi_p| \otimes \langle 0|)|\widetilde{\Psi_p}\rangle - H(p)| \le 2\epsilon/3$. By Theorem 8, we can use $\Theta(\ln(1/\beta)/\epsilon)$ applications of the unitaries (and their inverses) that implement $|\psi_p\rangle$ and $|\widetilde{\Psi_p}\rangle$ to estimate $(\langle\psi_p| \otimes \langle 0|)|\widetilde{\Psi_p}\rangle$ within additive error $\frac{\epsilon}{12\ln(2/\beta)}$. In total, this estimates $H(p)$ within additive error $\frac{\epsilon}{12\ln(2/\beta)} \cdot 4\ln(2/\beta) + \frac{2\epsilon}{3} = \epsilon$ with success probability at least $8/\pi^2$. The total complexity of the algorithm is

$$\mathcal{O}\left(\frac{\ln(1/\beta)}{\epsilon}\right) \cdot \mathcal{O}\left(\sqrt{\frac{n}{\epsilon}\log\left(\frac{n}{\epsilon}\right)}\log\left(\frac{\log n}{\epsilon}\right)\right) = \mathcal{O}\left(\frac{\sqrt{n}}{\epsilon^{1.5}}\log^{1.5}\left(\frac{n}{\epsilon}\right)\log\left(\frac{\log n}{\epsilon}\right)\right). \qquad \square$$

3.2 Density matrices with purified quantum query-access

For a density matrix ρ, we also assume the purified quantum query-access in Definition 2, i.e., a unitary oracle U_ρ acting as $U_\rho|0\rangle_A|0\rangle_B = |\rho\rangle = \sum_{i=1}^{n}\sqrt{p_i}|\phi_i\rangle_A|\psi_i\rangle_B$. We use U_ρ and $U_\rho^\dagger$ to estimate the von-Neumann entropy $H(\rho) = -\mathrm{Tr}[\rho\log\rho]$:

Theorem 13. *For any $0 < \epsilon < 1$, we can estimate $H(p)$ with accuracy ϵ with success probability at least $2/3$ using $\widetilde{\mathcal{O}}(\frac{n}{\epsilon^{1.5}})$ calls to U_ρ and $U_\rho^\dagger$.*

Proof. We use the construction of Eq. (2). The proof is essentially the same as that of Theorem 12 proceeding by constructing singular value transformation via Theorem 10, with the only difference that all probabilities are rescaled by a factor of $1/\sqrt{n}$ in (2); as a result, the number of calls to U_ρ and $U_\rho^\dagger$ is blown up to $\widetilde{\mathcal{O}}\left(\sqrt{n}\cdot\frac{\sqrt{n}}{\epsilon^{1.5}}\right) = \widetilde{\mathcal{O}}(\frac{n}{\epsilon^{1.5}})$. $\square$

4 Robust testers for ℓ^2-closeness with purified query-access

First we give an ℓ^2-closeness tester for unknown classical distributions p, q.

Theorem 14. *Given purified quantum query-access for classical distributions p, q as in Definition 2, for any $\nu, \epsilon \in (0, 1)$ the quantum query complexity of distinguishing the cases $\|p - q\|_2 \ge \epsilon$ and $\|p - q\|_2 \le (1 - \nu)\epsilon$ with success probability at least $2/3$ is $\mathcal{O}(\frac{1}{\nu\epsilon}\log^3(\frac{1}{\nu\epsilon})\log\log(\frac{1}{\nu\epsilon}))$.*

Proof. The main idea is to first bin the x elements based on the approximate value of $p(x) + q(x)$, then apply fine-tuned algorithms exploiting the knowledge of the approximate value of $p(x) + q(x)$.

Using amplitude estimation for any $k \in \mathbb{N}$ we can construct an algorithm $\mathcal{A}_k$ that for any input x with $p(x) + q(x) \geq 2^{-k}$ outputs "greater" with probability at least $2/3$, and for any x with $p(x) + q(x) \leq 2^{-k-1}$ outputs "smaller" and uses $\mathcal{O}\left(2^{\frac{k}{2}}\right)$ queries to U_p and U_q. Using $\mathcal{O}\big(\log(\frac{1}{\nu\epsilon})\big)$ repetitions we can boost the success probability to $1 - \mathcal{O}(\mathrm{poly}(\nu\epsilon))$. Since our algorithm only needs to succeed with constant probability, and will use these subroutines at most $\frac{1}{\mathrm{poly}(\nu\epsilon)}$ times, we can ignore the small failure probability. Therefore in the rest of the proof we assume without loss of generality, that $\mathcal{A}_k$ that solves perfectly the above question with (query) complexity $\mathcal{O}\left(2^{\frac{k}{2}} \log(\frac{1}{\nu\epsilon})\right)$.

Algorithm 1 Estimating $\log_2(p(x) + q(x))$

input $x \in [n]$, $\theta \in (0, 1)$
for $k \in K := \left\{-1, 0, 1, 2, \ldots, \left\lceil \log_2\left(\frac{1}{\theta}\right) \right\rceil\right\}$ **do**
    Run algorithm $\mathcal{A}_k$ on $|x\rangle$ **if** output is "greater" **then return** k
return "less than θ"

For any x with $p(x)+q(x) \geq \theta$, Algorithm 1 outputs a k such that $p(x)+q(x) \in (2^{-k-1}, 2^{-k+1})$. However, note that this labeling is probabilistic; let us denote by $s_k(x)$ the probability that x is labeled by k. Observe that $s_k(x) = 0$ unless $k \in \left\{ \left\lfloor \log_2\left(\frac{1}{p(x)+q(x)}\right)\right\rfloor, \left\lceil \log_2\left(\frac{1}{p(x)+q(x)}\right)\right\rceil \right\}$. Now let us express $\|p - q\|_2^2$ in terms of this "soft-selection" function $s(x)$.

$$\begin{aligned}
\|p - q\|_2^2 &= \sum_x |p(x) - q(x)|^2 \\
&= \sum_x \sum_{k \in K} s_k(x) |p(x) - q(x)|^2 + \eta && \eta \in [0, 2\theta) \\
&= \sum_{k \in K} 2^{9-k} \sum_x s_k(x) \frac{p(x) + q(x)}{2} \frac{2^{-k-2}}{p(x) + q(x)} \left(\frac{p(x) - q(x)}{2^{-k+3}}\right)^2 + \eta, && (10)
\end{aligned}$$

where the bound on η follows from the observation that

$$\eta \leq \sum_{x\colon p(x)+q(x)<\theta} |p(x) - q(x)|^2 \leq \sum_{x\colon p(x)+q(x)<\theta} (p(x) + q(x))^2 < \theta \sum_{x\colon p(x)+q(x)<\theta} p(x) + q(x) < 2\theta.$$

If for all $k \in K$ we have a $2^{k-9}\frac{\theta}{|K|}$-precise estimate of

$$\sum_x s_k(x) \frac{p(x) + q(x)}{2} \frac{2^{-k-2}}{p(x) + q(x)} \left(\frac{p(x) - q(x)}{2^{-k+3}}\right)^2, \tag{11}$$

then we get a 3θ-precise estimate of $\|p - q\|_2^2$. In particular setting $\theta := \nu\epsilon^2/6$, this solves the robust testing problem, since if $\|p - q\| \geq \epsilon$ then $\|p - q\|^2 \geq \epsilon^2$, on the other hand if $\|p - q\| \leq (1 - \nu)\epsilon$ then $\|p - q\|^2 \leq (1 - \nu)^2\epsilon^2 \leq (1 - \nu)\epsilon^2 = \epsilon^2 - \nu\epsilon^2$.

Now we describe how to construct a quantum algorithm that sets the first output qubit to $|0\rangle$ with probability (11). Start with preparing a purification of the distribution of $\frac{p(x)+q(x)}{2}$, then set the label of x to k with probability $s_k(x)$ using Algorithm 1 terminating it after using $\mathcal{A}_k$. Then separately apply the maps $\sqrt{\frac{2^{-k-2}}{p(x)+q(x)}}$ and $\frac{p(x)-q(x)}{2^{-k-3}}$ to the state.

Note that we do not need to apply the above transformations exactly, it is enough if apply them with precision say $2^{k-11}\frac{\theta}{|K|}$. We analyze the complexity of (approximately) implementing the above sketched algorithm. To implement the map $\sqrt{\frac{2^{-k-2}}{p(x)+q(x)}}$, we use the unitary of

Eq. (1), and transform the singular values by the polynomial $\tilde{P}$ from Corollary 11 using Theorem 10. In order to implement the map $\frac{p(x)-q(x)}{2^{-k-2}}$, we again use the unitary of Eq. (1), but now separately for p and q. We amplify both the singular values $\sqrt{p(x)}$ and $\sqrt{q(x)}$ by a factor $\sqrt{2^{k-2}}$ using the polynomial $\tilde{Q}$ from Corollary 11 in Theorem 10. Then we create a bolck-encoding[6] of both and $2^{k-2}p(x)$ and $2^{k-2}q(x)$ and then combine them to get a block-encoding of $\frac{p(x)-q(x)}{2^{-k-3}}$. In both cases the query complexity of $\mathcal{O}(\theta/|K|)$-precisely implementing the transformations is $\mathcal{O}\big(2^{k/2}\log(|K|/\theta)\big) = \mathcal{O}\big(2^{k/2}\log(1/\theta)\big)$. Since computing the label k also costs $\mathcal{O}\big(2^{k/2}\log(1/(\nu\epsilon))\big)$, this is the overall complexity so far. Finally we estimate the probability of the first qubit being set to $|0\rangle$ with setting $M = \mathcal{O}\big(|K|2^{-k/2}/(\nu\epsilon)\big)$ in Theorem 8, and boost the success probability to $1-\mathcal{O}(1/|K|)$ with $\mathcal{O}(\log(|K|))$ repetitions. Thus for any $k \in K$ the overall complexity of estimating Eq. (11) with sufficient precision has (query) complexity $\mathcal{O}\Big(\frac{|K|}{\nu\epsilon}\log\big(\frac{1}{\nu\epsilon}\big)\log(|K|)\Big) = \mathcal{O}\big(\frac{1}{\nu\epsilon}\log^2\big(\frac{1}{\nu\epsilon}\big)\log\log\big(\frac{1}{\nu\epsilon}\big)\big)$. Therefore estimating $\|p-q\|_2^2$ to precision $\nu\epsilon^2/6$ with high probability has (query) complexity

$$\mathcal{O}\left(\frac{1}{\nu\epsilon}\log^3\left(\frac{1}{\nu\epsilon}\right)\log\log\left(\frac{1}{\nu\epsilon}\right)\right). \qquad \square$$

It is easy to see an $\Omega\big(\frac{1}{\epsilon}\big)$ lower bound on the above problem even in the strongest quantum pure state input model Definition 4. Indeed, consider the case $n=2, q=(\frac{1}{2},\frac{1}{2})$ (the uniform distribution on $\{1,2\}$) and we want to test whether $p=q$ or $\|p-q\|_2 \geq \epsilon$. This is equivalent to test whether $p_1 = \frac{1}{2}$ or $|p_1 - \frac{1}{2}| \geq \frac{\epsilon}{\sqrt{2}}$; due to the optimality of amplitude estimation in Theorem 8, this task requires $\Omega(\frac{1}{\epsilon})$ quantum queries to the unitary U preparing the state $\sqrt{p_1}|1\rangle + \sqrt{p_2}|2\rangle$.

Now we prove the following result on (robust) ℓ^2-closeness testing for quantum distributions:

Theorem 15. *Given $\epsilon, \nu \in (0,1)$ and two density operators $\rho, \sigma \in \mathbb{C}^{n\times n}$ with purified quantum query-access to U_ρ and U_σ as in Definition 2, it takes $\mathcal{O}\Big(\min\Big(\frac{\sqrt{n}}{\epsilon}, \frac{1}{\epsilon^2}\Big)\frac{1}{\nu}\Big)$ queries to $U_\rho, U_\rho^\dagger, U_\sigma, U_\sigma^\dagger$ to decide whether $\|\rho-\sigma\|_2 \geq \epsilon$ or $\|\rho-\sigma\|_2 \leq (1-\nu)\epsilon$, with success probability at least $2/3$.*

Proof. We can combine the block-encodings of ρ and σ to apply the map $\frac{\rho-\sigma}{2}$ to the maximally entangled state $\sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}}$, which gives

$$\sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}} \to \left(\frac{\rho-\sigma}{2}\otimes I\right)\sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}}|0\rangle + \ldots|1\rangle.$$

The probability of measuring the $|0\rangle$ ancilla state is

$$\sum_{i,j=1}^n \frac{\langle i|\langle i|}{\sqrt{n}}\left(\frac{(\rho-\sigma)^2}{4}\otimes I\right)\frac{|j\rangle|j\rangle}{\sqrt{n}} = \frac{1}{4n}\sum_{i=1}^n \langle i|(\rho-\sigma)^2|i\rangle = \frac{1}{4n}\mathrm{Tr}[(\rho-\sigma)^2].$$

Thus it suffices to apply amplitude estimation with $M = \Theta\Big(\frac{\sqrt{n}}{\nu\epsilon}\Big)$ calls to $U_\rho, U_\rho^\dagger, U_\sigma, U_\sigma^\dagger$.

On the other hand, we can estimate $\|\rho-\sigma\|_2^2$ by observing that $\|\rho-\sigma\|_2^2 = \mathrm{Tr}\big[(\rho-\sigma)^2\big] = \mathrm{Tr}\big[\rho^2\big] - 2\mathrm{Tr}[\rho\sigma] + \mathrm{Tr}\big[\sigma^2\big]$. Since the success probability of the SWAP test (Buhrman et al. (2001)) on input states ρ, σ is $\frac{1}{2}(1+\mathrm{Tr}[\rho\sigma])$, we can individually estimate the latter quantities with precision $\mathcal{O}\big(\nu\epsilon^2\big)$ using amplitude estimation (Theorem 8) with $\mathcal{O}\big(\frac{1}{\nu\epsilon^2}\big)$ queries to $U_\rho, U_\rho^\dagger, U_\sigma, U_\sigma^\dagger$. As a result, we could decide whether $\|\rho-\sigma\|_2 \geq \epsilon$ or $\|\rho-\sigma\|_2 \leq (1-\nu)\epsilon$ using $\mathcal{O}\big(\frac{1}{\nu\epsilon^2}\big)$ queries.

The result of Theorem 15 hence follows by taking the minimum of the two complexities. $\square$

[6] If we have a projected unitary encoding of $\Pi U \widetilde{\Pi} = A = \sum_i \varsigma_i |\psi_i\rangle\langle 0,i|$ with $\widetilde{\Pi} = |0\rangle\langle 0| \otimes I$ we can immediately turn it into a block-encoding of $A^\dagger A = \sum_i \varsigma_i^2 |i\rangle\langle i|$, by e.g. applying Theorem 10 with the polynomial x^2.

5 Future work and open questions

Our paper raises a couple of natural open questions for future work. For example:

- Can we prove quantum lower bounds that match our upper bounds? For instance, can we prove an $\Omega\left(\frac{n}{\epsilon}\right)$ lower bound on estimating the von Neumann entropy in the purified quantum query-access model for density operators? Is there a lower bound technique which naturally fits our purified quantum query input model?
- For which other distributional property testing problems can we get speed-ups using the presented methodology?

Acknowledgments

A.G. thanks Ronald de Wolf, Ignacio Cirac and Yimin Ge for useful discussion.

References

Jayadev Acharya, Hirakendu Das, Alon Orlitsky, and Ananda Theertha Suresh. A unified maximum likelihood approach for estimating symmetric properties of discrete distributions. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 11–21, 2017a. URL http://proceedings.mlr.press/v70/acharya17a.html. arXiv: 1611.02960

Jayadev Acharya, Ibrahim Issa, Nirmal V. Shende, and Aaron B. Wagner. Measuring quantum entropy. arXiv: 1711.00814, 2017b.

Srinivasan Arunachalam and Ronald de Wolf. Optimal quantum sample complexity of learning algorithms. In *Proceedings of the 32nd IEEE Conference on Computational Complexity (CCC)*, pages 25:1–25:31, 2017. doi: 10.4230/LIPIcs.CCC.2017.25. arXiv: 1607.00932

Srinivasan Arunachalam, Sourav Chakraborty, Troy Lee, and Ronald de Wolf. Two new results about quantum exact learnings. arXiv: 1810.00481, 2018.

Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. arXiv: 1708.06002, 2017.

Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 442–451, 2001. doi: 10.1109/SFCS.2001.959920.

Tuğkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005. doi: 10.1145/509907.510005.

Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 60(1):4, 2013. doi: 10.1145/2432622.2432626. arXiv: 1009.5397

Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *Contemporary Mathematics Series*, pages 53–74. AMS, 2002. doi: 10.1090/conm/305. arXiv: quant-ph/0005055

Sergey Bravyi, Aram W. Harrow, and Avinatan Hassidim. Quantum algorithms for testing properties of distributions. *IEEE Transactions on Information Theory*, 57(6):3971–3981, 2011. doi: 10.1109/TIT.2011.2134250. arXiv: 0907.3920

Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001. doi: 10.1103/PhysRevLett.87.167902. arXiv: quant-ph/0102001

Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. In *Proceedings of the 50th ACM Symposium on Theory of Computing (STOC)*, 2018. doi: 10.1145/3188745.3188784. arXiv: 1710.09079

André Chailloux. A note on the quantum query complexity of permutation symmetric functions. In *Proceedings of the 10th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 19:1–19:7, 2018. doi: 10.4230/LIPIcs.ITCS.2019.19. arXiv: 1810.01790

Shantanav Chakraborty, András Gilyén, and Stacey Jeffery. The power of block-encoded matrix powers: improved regression techniques via faster Hamiltonian simulation. arXiv: 1804.01973, 2018.

Sourav Chakraborty, Eldar Fischer, Arie Matsliah, and Ronald de Wolf. New results on quantum property testing. In *Proceedings of the 30th International Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, page 145, 2010. doi: 10.4230/LIPIcs.FSTTCS.2010.145. arXiv: 1005.0523

Siu-On Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1193–1203, 2014. doi: 10.1137/1.9781611973402.88. arXiv: 1308.3946

Ilias Diakonikolas and Daniel M. Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 685–694, 2016. doi: 10.1109/FOCS.2016.78. arXiv: 1601.05557

András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. arXiv: 1806.01838, 2018.

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

Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Minimax estimation of functionals of discrete distributions. *IEEE Transactions on Information Theory*, 61(5):2835–2885, 2015. doi: 10.1109/TIT.2015.2412945. arXiv: 1406.6956

Tongyang Li and Xiaodi Wu. Quantum query complexity of entropy estimation. *IEEE Transactions on Information Theory*, pages 1–1, 2018. doi: 10.1109/TIT.2018.2883306. arXiv: 1710.06025

Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by qubitization. arXiv: 1610.06546, 2016.

Ashley Montanaro. Quantum speedup of Monte Carlo methods. *Proceedings of the Royal Society A*, 471(2181), 2015. doi: 10.1098/rspa.2015.0301. arXiv: 1504.06987

Ashley Montanaro and Ronald de Wolf. *A Survey of Quantum Property Testing*. Number 7 in Graduate Surveys. Theory of Computing Library, 2016. doi: 10.4086/toc.gs.2016.007. arXiv: 1310.2035

Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000. doi: 10.1017/CBO9780511976667.

Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the 47th ACM Symposium on Theory of Computing (STOC)*, pages 529–538, 2015. doi: 10.1145/2746539.2746582. arXiv: 1501.05028

Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC)*, pages 899–912, 2016. doi: 10.1145/2897518.2897544. arXiv: 1508.01907

Ryan O'Donnell and John Wright. Efficient quantum tomography II. In *Proceedings of the 49th ACM Symposium on Theory of Computing (STOC)*, pages 962–974, 2017. doi: 10.1145/3055399.3055454. arXiv: 1612.00034

Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6): 1191–1253, 2003. doi: 10.1162/089976603321780272.

Liam Paninski. Estimating entropy on m bins given fewer than m samples. *IEEE Transactions on Information Theory*, 50(9):2200–2203, 2004. doi: 10.1109/TIT.2004.833360.

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

Alfréd Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561. University of California Press, 1961. URL https://projecteuclid.org/euclid.bsmsp/1200512181.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Gregory Valiant and Paul Valiant. Estimating the unseen: an $n/log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pages 685–694, 2011a. doi: 10.1145/1993636.1993727.

Gregory Valiant and Paul Valiant. The power of linear estimators. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 403–412, 2011b. doi: 10.1109/FOCS.2011.81.

Yihong Wu and Pengkun Yang. Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Transactions on Information Theory*, 62(6):3702–3720, 2016. doi: 10.1109/TIT.2016.2548468. arXiv: 1407.0381

A Projected unitary encodings used for singular value transformation

First we handle the case of classical distributions. Let U_p be a purified quantum oracle of a classical distribution p as in Definition 2, and let $U := (U_p \otimes I)$, also let $\Pi := (\sum_{i=1}^n I \otimes |i\rangle\langle i| \otimes |i\rangle\langle i|)$, $\widetilde{\Pi} := (|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes I)$, then

$$\begin{aligned}
\Pi U \widetilde{\Pi} = \Pi (U_p \otimes I) \widetilde{\Pi} &= \Big(\sum_{i=1}^n I \otimes |i\rangle\langle i| \otimes |i\rangle\langle i| \Big) (U_p \otimes I) \big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes I \big) \\
&= \sum_{i=1}^n \Big((I \otimes |i\rangle\langle i|) U_p (|0\rangle\langle 0| \otimes |0\rangle\langle 0|) \Big) \otimes |i\rangle\langle i| \\
&= \sum_{i=1}^n \Big((I \otimes |i\rangle\langle i|) \sum_{j=1}^n \sqrt{p_j} |\phi_j\rangle |j\rangle \langle 0| \langle 0| \Big) \otimes |i\rangle\langle i| \\
&= \sum_{i=1}^n \sqrt{p_i} |\phi_i\rangle\langle 0| \otimes |i\rangle\langle 0| \otimes |i\rangle\langle i|.
\end{aligned}$$

Now we turn to quantum distributions where we do not know the diagonalizing basis of the density operator ρ. Let U_ρ be a purified quantum oracle of a quantum distribution ρ as in Definition 2, and W a unitary, mapping $|0\rangle|0\rangle \mapsto \sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}}$. Let $U' := \Big(I \otimes U_\rho^\dagger \Big) \big(W^\dagger \otimes I \big)$, $\Pi' := (I \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|)$ and $\widetilde{\Pi}$ as above, then

$$\begin{aligned}
\Pi' U' \widetilde{\Pi} = \Pi' \Big(I \otimes U_\rho^\dagger \Big) \Big(W^\dagger \otimes I \Big) \widetilde{\Pi} &= \Big(I \otimes (|0\rangle\langle 0| \otimes |0\rangle\langle 0| U_\rho^\dagger) \Big) \left(\left(\sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}} \right) \langle 0| \langle 0| \otimes I \right) \\
&= \left(I \otimes \sum_{i=1}^n \sqrt{p_i} |0\rangle |0\rangle \langle \phi_i | \langle \psi_i | \right) \left(\left(\sum_{j=1}^n \frac{|\phi'_j\rangle|\phi_j\rangle}{\sqrt{n}} \right) \langle 0| \langle 0| \otimes I \right) \\
&= \sum_{i=1}^n \sqrt{\frac{p_i}{n}} |\phi'_i\rangle |0\rangle |0\rangle \langle 0| \langle 0| \langle \psi_i|,
\end{aligned}$$

where $\sum_{j=1}^n \frac{|\phi'_j\rangle|\phi_j\rangle}{\sqrt{n}} = \sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}}$ is the Schmidt decomposition of the maximally entangled state under the basis $(|\phi_1\rangle, \ldots, |\phi_n\rangle)$.

B Polynomial approximations for singular value transformation

We use the following result based on local Taylor series:

Lemma 16. (Gilyén et al., 2018, Corollary 66) *Let $x_0 \in [-1, 1]$, $r \in (0, 2]$, $\nu \in (0, r]$ and let $f\colon [-x_0 - r - \nu, x_0 + r + \nu] \to \mathbb{C}$ and be such that $f(x_0 + x) = \sum_{\ell=0}^\infty a_\ell x^\ell$ for all $x \in [-r - \nu, r + \nu]$. Suppose $B > 0$ is such that $\sum_{\ell=0}^\infty (r + \nu)^\ell |a_\ell| \leq B$. Let $\epsilon \in \big(0, \frac{1}{2B}\big]$, then there is an efficiently computable polynomial $P \in \mathbb{C}[x]$ of degree $\mathcal{O}\big(\frac{1}{\nu} \log\big(\frac{B}{\epsilon}\big)\big)$ such that*[7]

$$\begin{aligned}
\|f(x) - P(x)\|_{[x_0 - r, x_0 + r]} &\leq \epsilon \\
\|P(x)\|_{[-1,1]} &\leq \epsilon + \|f(x)\|_{[x_0 - r - \nu/2, x_0 + r + \nu/2]} \leq \epsilon + B \\
\|P(x)\|_{[-1,1] \setminus [x_0 - r - \nu/2, x_0 + r + \nu/2]} &\leq \epsilon.
\end{aligned}$$

[7] For a function $g\colon \mathbb{R} \to \mathbb{C}$, and an interval $[a, b] \subseteq \mathbb{R}$, we define $\|g\|_{[a,b]} := \max_{x \in [a,b]} |g(x)|$.

We can use the above result to construct the following useful polynomial approximations.

Lemma 11. (Polynomial approximations) *Let $\beta \in (0,1]$, $\eta \in (0, \frac{1}{2}]$ and $t \geq 1$. There exists polynomials $\tilde{P}, \tilde{Q}, \tilde{S}$ such that*

- $\forall x \in [\frac{1}{t}, 1]$: $|\tilde{P}(x) - \frac{1}{2tx}| \leq \eta$, *and* $\forall x \in [-1,1]$: $-1 \leq \tilde{P}(x) = \tilde{P}(-x) \leq 1$,
- $\forall x \in [-\frac{1-\beta}{t}, \frac{1-\beta}{t}]$: $|\tilde{Q}(x) - tx| \leq \eta \cdot (tx)$, *and* $\forall x \in [-1,1]$: $\tilde{Q}(x) = -\tilde{Q}(-x) \leq 1$,
- $\forall x \in [\beta, 1]$: $|\tilde{S}(x) - \frac{\ln(1/x)}{2\ln(2/\beta)}| \leq \eta$, *and* $\forall x \in [-1,1]$: $-1 \leq \tilde{S}(x) = \tilde{S}(-x) \leq 1$,

moreover $\deg(\tilde{P}) = \mathcal{O}\Big(t \log\Big(\frac{1}{\eta}\Big)\Big)$, $\deg(\tilde{Q}) = \mathcal{O}\Big(\frac{t}{\beta} \log\Big(\frac{1}{\eta}\Big)\Big)$, *and* $\deg(\tilde{S}) = \mathcal{O}\Big(\frac{1}{\beta} \log\Big(\frac{1}{\eta}\Big)\Big)$.

Proof. For the construction of the $\tilde{P}$ and $\tilde{Q}$ polynomials see Corollary 67 and Theorem 30 of Gilyén et al. (2018), respectively. It remains to construct the polynomial $\tilde{S}$ above.

Denote $f(x) = \frac{\ln(1/x)}{2\ln(2/\beta)}$; by taking $\epsilon = \eta/2$, $x_0 = 1$, $r = 1-\beta$, $\nu = \frac{\beta}{2}$, and $B = \frac{1}{2}$ in Corollary 16, we have a polynomial $S \in \mathbb{C}[x]$ of degree $\mathcal{O}\big(\frac{1}{\nu}\log(\frac{B}{\epsilon})\big) = \mathcal{O}\Big(\frac{1}{\beta}\log(\frac{1}{\eta})\Big)$ such that

$$\|f(x) - S(x)\|_{[\beta, 2-\beta]} \leq \eta/2 \tag{12}$$

$$\|S(x)\|_{[-1,1]} \leq B + \eta/2 \leq (1+\eta)/2 \tag{13}$$

$$\|S(x)\|_{[-1, \frac{\beta}{2}]} \leq \eta/2. \tag{14}$$

Note that $B = \frac{1}{2}$ is valid because the local Taylor series of $f(x)$ at $x = 1$ is $\frac{1}{2\ln(2/\beta)} \sum_{l=1}^{\infty} \frac{(-1)^l x^l}{l}$, and as a result we could take

$$\begin{aligned} B = \frac{1}{2\ln(2/\beta)} \sum_{l=1}^{\infty} \frac{(1-\beta/2)^l}{l} &= -\frac{1}{2\ln(2/\beta)} \sum_{l=1}^{\infty} \frac{(-1)^{l-1}}{l} (-1+\beta/2)^l \\ &= -\frac{1}{2\ln(2/\beta)} \ln\frac{\beta}{2} = \frac{1}{2}. \end{aligned}$$

However, S is not an even polynomial in general; we instead take $\tilde{S}(x) = S(x) + S(-x)$ for all $x \in [-1,1]$. Then by (12) and (14) we have

$$\Big\|f(x) - \tilde{S}(x)\Big\|_{[\beta,1]} \leq \Big\|f(x) - \tilde{S}(x)\Big\|_{[\beta,1]} + \Big\|\tilde{S}(-x)\Big\|_{[\beta,1]} \leq \frac{\eta}{2} + \frac{\eta}{2} = \eta. \tag{15}$$

Furthermore, $\tilde{S}$ is an even polynomial such that $\deg(\tilde{S}) = \mathcal{O}\Big(\frac{1}{\beta}\log(\frac{1}{\eta})\Big)$; hence (13) and (14) imply

$$\Big\|\tilde{S}(x)\Big\|_{[-1,1]} = \Big\|\tilde{S}(x)\Big\|_{[0,1]} \leq \|S(x)\|_{[0,1]} + \|S(x)\|_{[-1,0]} \leq \frac{1+\eta}{2} + \frac{\eta}{2} \leq 1$$

given $\eta \leq 1/2$. (Finally we can take the real part of $\tilde{S}(x)$ if it has some complex coefficients.) □

C Corollaries of our ℓ^2-closeness testing results

C.1 ℓ^1-closeness testing with purified query-access

Corollary 17. *Given $\epsilon > 0$ and two distributions p, q on the domain $[n]$ with purified quantum query-access via U_p and U_q as in Definition 2, it takes $\widetilde{\mathcal{O}}\Big(\frac{\sqrt{n}}{\epsilon}\Big)$ queries to $U_p, U_p^\dagger, U_q, U_q^\dagger$ to decide whether $p = q$ or $\|p-q\|_1 \geq \epsilon$ with success probability at least $2/3$. Similarly for density operators $\rho, \sigma \in \mathbb{C}^{n\times n}$ with purified quantum query-access via U_ρ and U_σ, it takes $\mathcal{O}\big(\frac{n}{\epsilon}\big)$ queries to $U_\rho, U_\rho^\dagger, U_\sigma, U_\sigma^\dagger$ to decide whether $\rho = \sigma$ or $\|\rho-\sigma\|_1 \geq \epsilon$ with success probability at least $2/3$.*

Proof. By the Cauchy-Schwartz inequality we have $\|p-q\|_2 \geq \frac{1}{\sqrt{n}}\|p-q\|_1$, therefore Theorem 14 implies our claim by taking $\epsilon \leftarrow \epsilon/\sqrt{n}$ therein. Similarly, Theorem 15 implies our claim for quantum distributions ρ and σ. □

C.2 Independence testing with purified query-access

Corollary 18. *Given $\epsilon > 0$ and a classical distribution p on $[n] \times [m]$ with the purified quantum query-access via U_p as in Definition 2, it takes $\widetilde{\mathcal{O}}\left(\frac{\sqrt{nm}}{\epsilon}\right)$ queries to $U_p, U_p^\dagger$ to decide whether p is a product distribution on $[n] \times [m]$ or p is ϵ-far in ℓ^1-norm from any product distribution on $[n] \times [m]$ with success probability at least $2/3$.*

Proof. We define p_A to be the margin of p on the first marginal space, i.e., $p_A(i) = \sum_{j=1}^m p(i,j)$ for all $i \in [n]$. We similarly define p_B to be the margin of p on the second marginal space, i.e., $p_B(j) = \sum_{i=1}^n p(i,j)$ for all $j \in [m]$. Assume the quantum oracle U_p from Definition 2 acts as

$$U_p|0\rangle_A|0\rangle_B|0\rangle_C = \sum_{i=1}^n \sum_{j=1}^m \sqrt{p(i,j)}|i\rangle_A|j\rangle_B|\psi_{i,j}\rangle_C;$$

if we denote $|\phi_i\rangle = \sum_{j=1}^m \frac{\sqrt{p(i,j)}}{\sqrt{p_A(i)}}|j\rangle|\psi_{i,j}\rangle$ for all $i \in [n]$ and $|\varphi_j\rangle = \sum_{i=1}^n \frac{\sqrt{p(i,j)}}{\sqrt{p_B(j)}}|i\rangle|\psi_{i,j}\rangle$ for all $j \in [m]$, then we have

$$U_p|0\rangle_A|0\rangle_B|0\rangle_C = \sum_{i=1}^n \sqrt{p_A(i)}|i\rangle_A|\phi_i\rangle_{B,C} = \sum_{j=1}^m \sqrt{p_B(j)}|j\rangle_B|\varphi_j\rangle_{A,C}.$$

As a result,

$$(U_p \otimes U_p)(|0\rangle^{\otimes 6}) = \sum_{i=1}^n \sum_{j=1}^m \sqrt{p_A(i)}\sqrt{p_B(j)}|i\rangle|j\rangle|\phi_i\rangle|\varphi_j\rangle;$$

in other words, one purified quantum query to the distribution $p_A \times p_B$ can be implemented by two queries to U_p.

If p is a product distribution on $[n] \times [m]$, then $p = p_A \times p_B$; if p is ϵ-far in ℓ^1-norm from any product distribution on $[n] \times [m]$, then $\|p - p_A \times p_B\|_1 \geq \epsilon$. Therefore, the problem of independence testing reduces to ℓ^1-closeness testing for distributions on $[n] \times [m]$, and hence Corollary 18 follows from Corollary 17. □

Similarly, Corollary 17 implies that the quantum query complexity of testing independence of quantum distributions is $\mathcal{O}\left(\frac{nm}{\epsilon}\right)$.