Distributional Property Testing in a Quantum World

András Gilyén
QuSoft, CWI and University of Amsterdam, The Netherlands
gilyen@cwi.nl

Tongyang Li
Department of Computer Science, Institute for Advanced Computer Studies,
University of Maryland, USA
Joint Center for Quantum Information and Computer Science,
University of Maryland, USA
tongyang@cs.umd.edu

──── **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. We also introduce a novel access model for quantum distributions, enabling the coherent preparation of quantum samples, and propose a general framework that can naturally handle both classical and quantum distributions in a unified manner. Our framework generalizes and improves previous quantum algorithms for testing closeness between unknown distributions, testing independence between two distributions, and estimating the Shannon / von Neumann entropy of distributions. For classical distributions our algorithms significantly improve the precision dependence of some earlier results. We also show that in our framework procedures for classical distributions can be directly lifted to the more general case of quantum distributions, and thus obtain the first speed-ups for testing properties of density operators that can be accessed coherently rather than only via sampling.

1 Introduction

Distributional property testing is a fundamental problem in theoretical computer science (see, e.g. [22]) aiming at determining properties of probability distributions with the least number of independent samples. It 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 $[n] := \{1, \ldots, n\}$, it requires $\Theta(n/\epsilon^2)$ samples to "learn" p,

i.e., to construct a distribution q on $[n]$ such that $\|p - q\|_1 \le \epsilon$ 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 [19], 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 [30]) which focuses on investigating the quantum advantage in testing classical statistical properties. Quantum speed-ups have already been established for a few specific problems such as testing closeness between distributions [13, 29], testing identity to known distributions [18], estimating entropies [27], 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] (i.e., 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 independent 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 independent 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 only studied by a few earlier works [29, 24] and only in the special case of classical distributions. The motivation for this input model is the following: we can think about a density operator as the outcome of some physical process modeled by some black-box. Suppose that the black-box can generate samples on demand. Unlike in the classical (randomized) setting, in a quantum scenario in principle it is always possible to reverse every computational / physical process – including this black-box. If reversion is not feasible, then we get the plain sampling model; however if it is possible to reverse the (quantum) black-box then we get the purified query access model that we describe. For example, if a quantum computer produces the samples via, say, a Monte Carlo method, then the process is easily reversible. However, if the samples come from some source "outside the lab", then reversing the process might not be possible. Therefore, both input models (purified quantum query access and sampling access) are well-motivated. The surprising fact is that this subtle difference in the input models gives rise to significantly different complexities, as we show later for several problems.

---

[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 $\text{Tr}[\rho] = 1$. Please refer to the textbook [31] for more information.

▶ **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]*

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

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., [13, 18, 27, 15].

▶ **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 $f: S \to [n]$ 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*

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

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 is when we assume access to a unitary (and its inverse) acting as $|0\rangle \mapsto \sum_{i=1}^{n} \sqrt{p_i}|i\rangle$. A very similar input model was thoroughly studied by [4].

▶ **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*

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

This is strictly stronger[3] 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

---

[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 ith 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.

[3] This can be seen in various ways. We give an argument in the spirit of distributional property testing. Closeness of two unknown distributions p, q can be tolerantly tested in the squared Hellinger distance $H(p, q)^2 = \frac{1}{2}\left\|\sqrt{p} - \sqrt{q}\right\|_2^2$ to precision ϵ in query complexity $\mathcal{O}\left(1/\sqrt{\epsilon}\right)$ in the model of Definition 4 using amplitude estimation. On the other hand the classical sample complexity of testing equality to ϵ precision in this metric is $\widetilde{\Theta}(\min(n^{2/3}/\epsilon^{4/3}, n^{3/4}/\epsilon))$, as shown in [20]. The results of Chailloux [16] imply that this query complexity improves at most cubically in the model of Definition 2, showing that the input model of Definition 4 is strictly stronger.

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 [6, 5].

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. In the classical literature these are well-studied properties, and the corresponding testers motivate general algorithms for testing properties of discrete distributions [20, 1].

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 $\frac{2}{3}$. Tolerant 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 $H(\rho) = -\text{Tr}[\rho \log(\rho)]$ within additive ϵ-precision with success probability at least $\frac{2}{3}$.*

1.2 New results

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,[4] as we prove in Theorem 12 and Theorem 13.
- Tolerant 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.)

---

[4] Throughout the paper, we use $\widetilde{\mathcal{O}}$ to omit poly-logarithmic factors in $\mathcal{O}$, i.e., $\widetilde{\mathcal{O}}(T) = \mathcal{O}(T \operatorname{poly}(\log T))$.
[5] Recent results of [16] imply that in this model quantum speed-ups are at most cubic.

▪ **Table 1** Summary of sample and query complexity results. Our new bounds are printed in **bold**. For classical distributions with quantum query-access[5] we prove (almost) matching upper and lower bounds for ℓ^2-testing, and improve the previous best complexity $\widetilde{\mathcal{O}}\big(\sqrt{n}/\epsilon^{2.5}\big)$ for ℓ^1-testing by [29] and $\widetilde{\mathcal{O}}\big(\sqrt{n}/\epsilon^2\big)$ for Shannon entropy estimation by [27]. We are not aware of prior work on testing quantum distributions with purified query-access.

model \ problem	ℓ^1-closeness testing	(tolerant) ℓ^2-closeness testing	Shannon / von Neumann entropy
Classical distribution sampling	$\Theta\left(\max\left\{\frac{n^{2/3}}{\epsilon^{4/3}}, \frac{n^{1/2}}{\epsilon^2}\right\}\right)$ [19]	$\Theta\big(\frac{1}{\epsilon^2}\big)$ [19]	$\Theta\left(\frac{n}{\epsilon \log n} + \frac{\log^2 n}{\epsilon^2}\right)$ [25, 42]
Classical distribution with purified query-access	$\widetilde{\mathcal{O}}\big(\frac{\sqrt{n}}{\epsilon}\big)$	$\widetilde{\Theta}\big(\frac{1}{\epsilon}\big)$	$\widetilde{\mathcal{O}}\big(\frac{\sqrt{n}}{\epsilon^{1.5}}\big)$; $\widetilde{\Omega}(\sqrt{n})$ [15]
Quantum state with purified query-access	$\mathcal{O}\big(\frac{n}{\epsilon}\big)$	$\mathcal{O}\left(\min\left(\frac{\sqrt{n}}{\epsilon}, \frac{1}{\epsilon^2}\right)\right)$	$\widetilde{\mathcal{O}}\big(\frac{n}{\epsilon^{1.5}}\big)$
Quantum state sampling	$\Theta\big(\frac{n}{\epsilon^2}\big)$ [7]	$\Theta\big(\frac{1}{\epsilon^2}\big)$ [7]	$\mathcal{O}\big(\frac{n^2}{\epsilon^2}\big), \Omega\big(\frac{n^2}{\epsilon}\big)$ [3]

As we show our quantum algorithms for classical distributional property testing problems with purified access can be naturally lifted to the case of quantum distributions, incurring an overhead of $\approx \sqrt{n}$, which is manifested in the complexities of Table 1.

▪ **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.

	Sample complexity	(Purified) Query complexity
Classical	$\Theta\big(\frac{n}{\log n}\big)$ [40]	$\widetilde{\Theta}\big(\sqrt{n}\big)$ [27, 15]
Quantum	$\Theta\big(n^2\big)$ [3]	$\widetilde{\mathcal{O}}(n)$

1.3 New 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 due to [28]. We call such a unitary a *block-encoding* of ρ:

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

One can think of a block-encoding as a post-selective 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 [17].

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 we can also prepare a purification of $(\rho + \sigma)/2$ by setting a qubit to the state

$(|0\rangle + |1\rangle)/\sqrt{2}$, and then controlled on the $|0\rangle$ or $|1\rangle$ value of the qubit run the process that samples from ρ or σ, respectively. 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.

One of our main technical contributions is to use a new type of block-encoding, which is a unitary matrix having a certain block proportional to a matrix A such that $A^\dagger A = \rho$, i.e., we use a "square-root" of the (quantum) distribution ρ (in the case of classical distributions rho is a diagonal matrix with the probabilities as diagonal entries). This new technique allows us to develop a unified approach for distributional property testing, which we consider one of our major contributions. It is this new perspective that enables us to derive several results in a relatively short paper. Once we establish this methodology the results are relatively easy to derive in a systematic way, both for classical and quantum distributions.

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 a slight generalization of block-encodings called *projected unitary encodings*, which represent 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 post-selective 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 diagonalizing basis of ρ. Therefore we use slightly different operators. Let W be a unitary,[6] mapping $|0\rangle|0\rangle \mapsto \sum_{j=1}^n \frac{|j\rangle|j\rangle}{\sqrt{n}}$. Let $U' := \big(I \otimes U_\rho^\dagger\big)(W^\dagger \otimes I)$, $\Pi' := (I \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|)$ and $\widetilde{\Pi}$ as above. As we show in Appendix A these operators form a projected unitary encoding of

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

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 $(|\phi_1\rangle, \ldots, |\phi_n\rangle)$.

---

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

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. We note that for (approximately) transforming a block-encoding of $A/\sqrt{n}$ to a block-encoding of $A/\mathcal{O}(1)$ it is necessary and sufficient to use the block-encoding of $A/\sqrt{n}$ about $\sqrt{n}$ times [21, Theorems 3&17]. This is essentially the reason for the $\approx \sqrt{n}$ overhead in our quantum algorithms for quantum distributions in Table 1. 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. This suggests that it is impossible to obtain a more efficient block-encoding in the general case.

General recipe. We summarize our recipe to distributional property testing as follows.

1. Construct quantum circuit / unitary matrix 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 tolerant 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 $\|p - q\|_1 \geq \epsilon$, [10] first gave a sublinear algorithm using $\tilde{O}(n^{2/3}/\epsilon^{8/3})$ samples to p and q. The follow-up work by [19] 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 [9] to be $\tilde{O}(\sqrt{n}/\epsilon^4)$ and then [37] gave the tight bound $\Theta(\sqrt{n}/\epsilon^2)$. More recently, Ref. [20] 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* (see also the previous studies by [9, 26, 2]), 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* [39] because it measures for example compressibility. The sample complexity of estimating $H(p)$ within additive error ϵ has been intensively studied [8, 35, 36]; in particular, [40, 41] 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 [25] and [42] gave the optimal estimator with $\Theta\left(\frac{n}{\epsilon \log n} + \frac{(\log n)^2}{\epsilon^2}\right)$ samples.

Quantum algorithms. The first paper on distributional property testing by quantum algorithms was by [13], 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, [18] gave an algorithm for identity testing (to an arbitrary known distribution) with $\widetilde{\mathcal{O}}(n^{1/3}/\epsilon^5)$ queries, and [29] improved the ϵ-dependence of ℓ^1-closeness testing to $\widetilde{\mathcal{O}}(\sqrt{n}/\epsilon^{2.5})$. More recently, [27] studied entropy estimation under this model, and gave a quantum algorithm for Shannon entropy estimation with $\widetilde{\mathcal{O}}(\sqrt{n}/\epsilon^2)$ queries and also sublinear quantum algorithms for estimating Rényi entropies ([38]).

Another type of quantum property testing results ([32, 33, 34, 3, 7, 23]) 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 $\Theta(1/\epsilon^2)$ ([7]), respectively. In addition [3] gave upper and lower bounds $\mathcal{O}(n^2/\epsilon^2), \Omega(n^2/\epsilon)$ 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 tolerantly testing ℓ^2-closeness of classical distributions, and another efficient tolerant 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** ([12, 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, [21] proposed *quantum singular value transformation* which turns out 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 transform *of* 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** ([21, 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*

$$P^{(SV)}\left(\widetilde{\Pi}U\Pi\right) = \begin{cases} \left(\langle+| \otimes \widetilde{\Pi}\right)\left(|0\rangle\langle0|\otimes U_\Phi + |1\rangle\langle1|\otimes U_{-\Phi}\right)\left(|+\rangle \otimes \Pi\right) & \text{if } n \text{ is odd, and} \\ \left(\langle+| \otimes \Pi\right)\left(|0\rangle\langle0|\otimes U_\Phi + |1\rangle\langle1|\otimes U_{-\Phi}\right)\left(|+\rangle \otimes \Pi\right) & \text{if } n \text{ is even,} \end{cases}$$

where $U_\Phi = e^{i\phi_1(2\widetilde{\Pi}-I)}U \prod_{j=1}^{(n-1)/2}\left(e^{i\phi_{2j}(2\Pi-I)}U^\dagger e^{i\phi_{2j+1}(2\widetilde{\Pi}-I)}U\right)$.[7]

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.

---

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

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

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 $\mathcal{O}\left(\frac{\sqrt{n}}{\epsilon^{1.5}} \log^{1.5}\left(\frac{n}{\epsilon}\right) \log\left(\frac{\log n}{\epsilon}\right)\right)$ calls to U_p and $U_p^\dagger$.*

Proof. The general idea is to first construct a unitary matrix that has a specific matrix block 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 Lemma 11, setting $\eta = \frac{\epsilon}{24 \ln(2/\beta)}$ and $\beta = \sqrt{\Delta}$ for $\Delta = \frac{\epsilon}{12n \ln(n/\epsilon)}$. Notice that this Δ satisfies

$$\Delta\left(\ln\left(\frac{1}{\Delta}\right) + 4\ln\left(\frac{2}{\beta}\right)\right) = \frac{\epsilon}{12n \ln(n/\epsilon)} \cdot \ln\frac{16(4n \ln\frac{n}{\epsilon})^3}{\epsilon^3} \leq \frac{\epsilon}{12n \ln(n/\epsilon)} \cdot \ln\frac{n^6}{\epsilon^6} = \frac{\epsilon}{2n}, \tag{6}$$

provided that $\frac{n}{\epsilon} \geq 152$. 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 $|\psi_p\rangle$ gives

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

Preparing $|\widetilde{\Psi_p}\rangle$ costs $\deg \tilde{S} = \mathcal{O}\left(\frac{1}{\beta} \log\left(\frac{1}{\eta}\right)\right) = \mathcal{O}\left(\sqrt{\frac{n}{\epsilon} \log\left(\frac{n}{\epsilon}\right)} \log\left(\frac{\log n}{\epsilon}\right)\right)$ uses of U_p and $U_p^\dagger$ and the same number of controlled reflections through $\Pi, \tilde{\Pi}$. Furthermore, Lemma 11 implies that for all i such that $p_i \geq \Delta$,

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

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

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

where the first inequality comes from the fact that $|\tilde{S}(x)| \leq 1$ for all $x \in [-1, 1]$, the second inequality comes from the monotonicity of $x(\ln(1/x) + 4\ln(2/\beta))$ on $(0, \frac{1}{\Delta}]$, and the third inequality comes from (6). As a result of (5), (8), and (9), we have

$$\left| (\langle\psi_p| \otimes \langle0|)|\widetilde{\Psi}_p\rangle - \frac{H(p)}{4\ln(2/\beta)} \right| = \left| p_i \tilde{S}(\sqrt{p_i}) - \sum_{i=1}^{n} \frac{p_i \log(1/p_i)}{4\ln(2/\beta)} \right|$$

$$\leq \sum_{i:\, p_i < \Delta} \frac{\epsilon}{8n\ln(2/\beta)} + \sum_{i:\, p_i \geq \Delta} \eta p_i$$

$$\leq \frac{\epsilon}{8\ln(2/\beta)} + \frac{\epsilon}{24\ln(2/\beta)} = \frac{\epsilon}{6\ln(2/\beta)}.$$

Therefore, $|4\ln(2/\beta)(\langle\psi_p| \otimes \langle0|)|\widetilde{\Psi}_p\rangle - H(p)| \leq 2\epsilon/3$. 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 $(\langle\psi_p| \otimes \langle0|)|\widetilde{\Psi}_p\rangle$ 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

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

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 $H(\rho) = -\operatorname{Tr}[\rho\log\rho]$:

▶ **Theorem 13.** *For any $0 < \epsilon < 1$, we can estimate $H(p)$ with accuracy ϵ with success probability at least $2/3$ using $\widetilde{\mathcal{O}}\left(\frac{n}{\epsilon^{1.5}}\right)$ 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 $\widetilde{\mathcal{O}}\left(\sqrt{n} \cdot \frac{\sqrt{n}}{\epsilon^{1.5}}\right) = \widetilde{\mathcal{O}}\left(\frac{n}{\epsilon^{1.5}}\right)$. ◀

4 Tolerant 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 $\|p - q\|_2 \geq \epsilon$ and $\|p - q\|_2 \leq (1 - \nu)\epsilon$ with success probability at least $2/3$ is*

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

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 $\mathcal{O}\left(\log(\frac{1}{\nu\epsilon})\right)$ repetitions we can boost the success probability to $1 - \mathcal{O}(\operatorname{poly}(\nu\epsilon))$. Since our

algorithm only needs to succeed with constant probability, and will use these subroutines at most $\frac{1}{\text{poly}(\nu\epsilon)}$ 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 $x \in [n]$, $\theta \in (0,1)$
1: **for** $k \in K := \left\{-1, 0, 1, 2, \ldots, \left\lceil \log_2\left(\frac{1}{\theta}\right)\right\rceil\right\}$ **do**
2:     Run algorithm $\mathcal{A}_k$ on $|x\rangle$ **if** output is "greater" **then return** k
3: **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 $k \in \left\{\left\lfloor\log_2\left(\frac{1}{p(x)+q(x)}\right)\right\rfloor, \left\lceil\log_2\left(\frac{1}{p(x)+q(x)}\right)\right\rceil\right\}$ (otherwise the return is either "greater" or "less than"). 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

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

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 tolerant 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 Lemma 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 Lemma 11 in Theorem 10. Then we create a block-encoding[8] 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 $\mathcal{O}\bigl(2^{k/2}\log(|K|/\theta)\bigr) = \mathcal{O}\bigl(2^{k/2}\log(1/\theta)\bigr)$. Since computing the label k also costs $\mathcal{O}\bigl(2^{k/2}\log(1/(\nu\epsilon))\bigr)$, this is the overall complexity so far. Finally we estimate the probability of the first qubit being set to $|0\rangle$ with setting $M = \mathcal{O}\bigl(|K|2^{-k/2}/(\nu\epsilon)\bigr)$ 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 $\mathcal{O}\bigl(\frac{|K|}{\nu\epsilon}\log\bigl(\frac{1}{\nu\epsilon}\bigr)\log(|K|)\bigr) = \mathcal{O}\bigl(\frac{1}{\nu\epsilon}\log^2\bigl(\frac{1}{\nu\epsilon}\bigr)\log\log\bigl(\frac{1}{\nu\epsilon}\bigr)\bigr)$. Therefore estimating $\|p-q\|_2^2$ to precision $\nu\epsilon^2/6$ with high probability has (query) complexity

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

It is easy to see an $\Omega\bigl(\frac{1}{\epsilon}\bigr)$ 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 result below on (tolerant) 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 $\mathcal{O}\left(\min\left(\frac{\sqrt{n}}{\epsilon}, \frac{1}{\epsilon^2}\right)\frac{1}{\nu}\right)$ 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 $M = \Theta\left(\frac{\sqrt{n}}{\nu\epsilon}\right)$ 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 $\|\rho - \sigma\|_2^2 = \mathrm{Tr}\bigl[(\rho-\sigma)^2\bigr] = \mathrm{Tr}\bigl[\rho^2\bigr] - 2\mathrm{Tr}[\rho\sigma] + \mathrm{Tr}\bigl[\sigma^2\bigr]$. Since the success probability of the SWAP test ([14]) on input states ρ, σ is $\frac{1}{2}(1 + \mathrm{Tr}[\rho\sigma])$, we can individually estimate the latter quantities with precision $\mathcal{O}(\nu\epsilon^2)$ using amplitude estimation (Theorem 8) with $\mathcal{O}\bigl(\frac{1}{\nu\epsilon^2}\bigr)$ 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 $\mathcal{O}\bigl(\frac{1}{\nu\epsilon^2}\bigr)$ queries.

The result of Theorem 15 hence follows by taking the minimum of the two complexities. ◀

---

[8] 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, including:

- For which other distributional property testing problems can we get faster and simpler quantum algorithms using the presented methodology?
- 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?
- Can we prove the conjecture that the purified and discrete query input models are equivalent for classical distributions, with respect to the query complexity of (distributional) property testing problems? For some recent progress in this direction see [11].

References

1 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, 2017. URL: `http://proceedings.mlr.press/v70/acharya17a.html`, `arXiv:1611.02960`.

2 Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pages 3591–3599, 2015. URL: `http://papers.nips.cc/paper/5839-optimal-testing-for-properties-of-distributions`, `arXiv:1507.05952`.

3 Jayadev Acharya, Ibrahim Issa, Nirmal V. Shende, and Aaron B. Wagner. Measuring Quantum Entropy. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 3012–3016, 2019. `arXiv:1711.00814`.

4 Dorit Aharonov and Amnon Ta-Shma. Adiabatic Quantum State Generation. *SIAM Journal on Computing*, 37(1):47–82, 2007. Earlier version in STOC'03, `arXiv:quant-ph/0301023`. `doi:10.1137/060648829`.

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

6 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`.

7 Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. In *Proceedings of the 51st ACM Symposium on the Theory of Computing (STOC)*, pages 503–514. ACM, 2019. `doi:10.1145/3313276.3316344`.

8 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`.

9 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`.

10 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`.

11 Aleksandrs Belovs. Quantum Algorithms for Classical Probability Distributions, 2019. `arXiv:1904.02192`.

12 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`.

13 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`.

14 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`.

15 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 the Theory of Computing (STOC)*, 2018. `doi:10.1145/3188745.3188784`.

16 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`.

17 Shantanav Chakraborty, András Gilyén, and Stacey Jeffery. The power of block-encoded matrix powers: improved regression techniques via faster Hamiltonian simulation. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*, 2019. `doi:10.4230/LIPIcs.ICALP.2019.28`.

18 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`.

19 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`.

20 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`.

21 András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st ACM Symposium on the Theory of Computing (STOC)*, 2019. `doi:10.1145/3313276.3316366`.

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

23 Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. `arXiv:1508.01797`.

24 Yassine Hamoudi and Frédéric Magniez. Quantum Chebyshev's Inequality and Applications. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 69:1–69:16, 2019. `doi:10.4230/LIPIcs.ICALP.2019.69`.

25 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`.

26 Reut Levi, Dana Ron, and Ronitt Rubinfeld. Testing properties of collections of distributions. *Theory of Computing*, 9(1):295–347, 2013. `doi:10.4086/toc.2013.v009a008`.

27 Tongyang Li and Xiaodi Wu. Quantum Query Complexity of Entropy Estimation. *IEEE Transactions on Information Theory*, 65(5):2899–2921, 2019. `doi:10.1109/TIT.2018.2883306`.

28 Guang Hao Low and Isaac L. Chuang. Hamiltonian Simulation by Qubitization. *Quantum*, 3:163, 2019. `doi:10.22331/q-2019-07-12-163`.

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

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

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

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

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

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

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

36   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`.

37   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`.

38   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`.

39   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`.

40   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 the Theory of Computing (STOC)*, pages 685–694, 2011. `doi:10.1145/1993636.1993727`.

41   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, 2011. `doi:10.1109/FOCS.2011.81`.

42   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`.

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 $\Pi := \left(\sum_{i=1}^n I \otimes |i\rangle\langle i| \otimes |i\rangle\langle i| \right)$, $\widetilde{\Pi} := (|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes I)$, then

$$
\Pi U \widetilde{\Pi} = \Pi (U_p \otimes I) \widetilde{\Pi} = \Big(\sum_{i=1}^n I \otimes |i\rangle\langle i| \otimes |i\rangle\langle i| \Big) (U_p \otimes I) \big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes I \big)
$$

$$
= \sum_{i=1}^n \Big((I \otimes |i\rangle\langle i|) U_p (|0\rangle\langle 0| \otimes |0\rangle\langle 0|) \Big) \otimes |i\rangle\langle i|
$$

$$\ldots = \sum_{i=1}^{n} \left((I \otimes |i\rangle\langle i|) \sum_{j=1}^{n} \sqrt{p_j} |\phi_j\rangle |j\rangle \langle 0|\langle 0| \right) \otimes |i\rangle\langle i|$$

$$= \sum_{i=1}^{n} \sqrt{p_i} |\phi_i\rangle\langle 0| \otimes |i\rangle\langle 0| \otimes |i\rangle\langle i|.$$

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 $U' := (I \otimes U_\rho^\dagger)(W^\dagger \otimes I)$, $\Pi' := (I \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|)$ and $\widetilde{\Pi}$ as above, then

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

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** ([21, Corollary 66]). *Let $x_0 \in [-1, 1]$, $r \in (0, 2]$, $\nu \in (0, r]$ and let $f : [-x_0 - r - \nu, x_0 + r + \nu] \to \mathbb{C}$ 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 $\sum_{\ell=0}^{\infty} (r + \nu)^\ell |a_\ell| \le B$. Let $\epsilon \in \left(0, \frac{1}{2B}\right]$, then there is an efficiently computable polynomial $P \in \mathbb{C}[x]$ of degree $\mathcal{O}\left(\frac{1}{\nu} \log\left(\frac{B}{\epsilon}\right)\right)$ such that[9]*

$$\|f(x) - P(x)\|_{[x_0-r, x_0+r]} \le \epsilon$$

$$\|P(x)\|_{[-1,1]} \le \epsilon + \|f(x)\|_{[x_0-r-\nu/2, x_0+r+\nu/2]} \le \epsilon + B$$

$$\|P(x)\|_{[-1,1]\setminus[x_0-r-\nu/2, x_0+r+\nu/2]} \le \epsilon.$$

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 $t \ge 1$. There exists polynomials $\tilde{P}, \tilde{Q}, \tilde{S}$ such that*
- *$\forall x \in [\frac{1}{t}, 1]$: $|\tilde{P}(x) - \frac{1}{2tx}| \le \eta$, and $\forall x \in [-1, 1]$: $-1 \le \tilde{P}(x) = \tilde{P}(-x) \le 1$,*
- *$\forall x \in [-\frac{1-\beta}{t}, \frac{1-\beta}{t}]$: $|\tilde{Q}(x) - tx| \le \eta \cdot (tx)$, and $\forall x \in [-1, 1]$: $\tilde{Q}(x) = -\tilde{Q}(-x) \le 1$,*
- *$\forall x \in [\beta, 1]$: $|\tilde{S}(x) - \frac{\ln(1/x)}{2\ln(2/\beta)}| \le \eta$, and $\forall x \in [-1, 1]$: $-1 \le \tilde{S}(x) = \tilde{S}(-x) \le 1$,*

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

---

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

Proof. For the construction of the $\tilde{P}$ and $\tilde{Q}$ polynomials see Corollary 67 and Theorem 30 of [21], 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 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

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

$$= -\frac{1}{2\ln(2/\beta)}\ln\frac{\beta}{2} = \frac{1}{2}.$$

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

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

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

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

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 $\mathcal{O}\left(\frac{nm}{\epsilon}\right)$.