Improved Sample Upper and Lower Bounds for Trace Estimation of Quantum State Powers

Kean Chen KEANCHEN.GAN@GMAIL.COM
Department of Computer and Information Science, University of Pennsylvania

Qisheng Wang QISHENGWANG1994@GMAIL.COM
School of Informatics, University of Edinburgh

Editors: Nika Haghtalab and Ankur Moitra

Abstract

As often emerges in various basic quantum properties such as entropy, the trace of quantum state powers $\mathrm{tr}(\rho^q)$ has attracted a lot of attention. The recent work of Liu and Wang (SODA 2025) showed that $\mathrm{tr}(\rho^q)$ can be estimated to within additive error ε with a dimension-independent sample complexity of $\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$ for any constant $q > 1$,[1] where only an $\Omega(1/\varepsilon)$ lower bound was given. In this paper, we significantly improve the sample complexity of estimating $\mathrm{tr}(\rho^q)$ in both the upper and lower bounds. In particular:

- For $q > 2$, we settle the sample complexity with matching upper and lower bounds $\widetilde{\Theta}(1/\varepsilon^2)$.
- For $1 < q < 2$, we provide an upper bound $\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$, with a lower bound $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1},2\}})$ for dimension-independent estimators, implying there is only room for a quadratic improvement.

Our upper bounds are obtained by (non-plug-in) quantum estimators based on weak Schur sampling, in sharp contrast to the prior approach based on quantum singular value transformation and samplizer.

Keywords: Quantum computing, sample complexity, trace estimation, sample lower bounds.

1. Introduction

Testing the properties of quantum states is a fundamental problem in the field of quantum property testing Montanaro and de Wolf (2016), where the spectra of quantum states turn out to be crucial, as they fully characterize unitarily invariant properties. Given samples of the quantum state to be tested, in O'Donnell and Wright (2021), testing the spectrum was extensively studied, with several significant applications such as mixedness testing and rank testing. In O'Donnell and Wright (2017), they further investigated the sample complexity of the spectrum tomography of quantum states. Subsequently, as a representative unitarily invariant quantity, the entropy of a quantum state was known to have efficient estimators in Acharya et al. (2020); Bavarian et al. (2016); Wang and Zhang (2024a).

The traces of quantum state powers, $\mathrm{tr}(\rho^q)$, of a quantum state ρ are one of the simplest functionals of quantum states. The quantity $\mathrm{tr}(\rho^q)$ has connections to the Rényi entropy $\mathrm{S}_q^{\mathrm{R}}(\rho) = \frac{1}{1-q}\ln(\mathrm{tr}(\rho^q))$ Rényi (1961) and the Tsallis entropy $\mathrm{S}_q^{\mathrm{T}}(\rho) = \frac{1}{1-q}(\mathrm{tr}(\rho^q) - 1)$ Tsallis (1988). The estimation of $\mathrm{tr}(\rho^q)$ is at the core of Tsallis entropy estimation, with a wide range of applications in physics. A notable example is the Tsallis entropy of order $q = \frac{3}{2}$ for modeling fluid dynamics

1. Throughout this paper, $\widetilde{O}(\cdot)$, $\widetilde{\Omega}(\cdot)$, and $\widetilde{\Theta}(\cdot)$ suppress polylogarithmic factors in ε.

systems Beck (2001, 2002). In addition, for $q = 1.001$ (close to 1), the Tsallis entropy $\mathrm{S}_q^{\mathrm{T}}(\rho)$ serves as a lower bound on the von Neumann entropy, whereas the former can be estimated exponentially faster than the latter, as noted in Liu and Wang (2025). In particular, $\mathrm{tr}(\rho^2)$ refers to the purity of ρ, and it is well-known that the purity $\mathrm{tr}(\rho^2)$ can be estimated to within additive error using $O(1/\varepsilon^2)$ samples of ρ via the SWAP test Buhrman et al. (2001). For the case of constant integer $q \geq 2$, $\mathrm{tr}(\rho^q)$ can be estimated using $O(1/\varepsilon^2)$ samples of ρ via the Shift test proposed in Ekert et al. (2002), generalizing the SWAP test. For non-integer $q > 0$ and $q \neq 1$, the estimation of $\mathrm{tr}(\rho^q)$ was considered in Wang et al. (2024a) with the corresponding quantum algorithms presented with time complexity $\mathrm{poly}(r, 1/\varepsilon)$,[2] where r is the rank of ρ. Recently in Liu and Wang (2025), it was discovered that for every non-integer $q > 1$, $\mathrm{tr}(\rho^q)$ can be estimated using $\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$ samples of ρ, removing the dependence on r (which we call dimension-independent as it depends on neither the rank nor the dimension of ρ). Thus, this exponentially improving the results in Wang et al. (2024a) and the results implied by other works Acharya et al. (2020); Wang et al. (2024b); Wang and Zhang (2024a) on Rényi entropy estimation. However, the sample complexity in Liu and Wang (2025) is far from being optimal, as only a lower bound of $\Omega(1/\varepsilon)$ on the sample complexity of estimating $\mathrm{tr}(\rho^q)$ for non-integer $q > 1$ was known in (Liu and Wang, 2025, Theorem 5.9). To our knowledge, only a matching lower bound of $\Omega(1/\varepsilon^2)$ was known for the case of $q = 2$, i.e., estimating the purity $\mathrm{tr}(\rho^2)$ (see (Chen et al., 2023, Theorem 5) and (Gong et al., 2024, Lemma 3)).

In this paper, we further investigate the sample complexity of estimating $\mathrm{tr}(\rho^q)$ for non-integer $q > 1$, achieving significant improvements over the prior results in both the upper and lower bounds. In particular, for $q > 2$, we provide an estimator that is *optimal* only up to a logarithmic factor in the precision ε. Our results are collected in Section 1.1. In addition, it is noteworthy that our techniques are conceptually and technically different from those in Liu and Wang (2025). In comparison, our estimator is based on weak Schur sampling Childs et al. (2007) while the estimator in Liu and Wang (2025) is based on quantum singular value transformation Gilyén et al. (2019) and samplizer Wang and Zhang (2023, 2024a). For more details, see Section 1.2.

1.1. Main Results

To illustrate our results, we present them in two parts separately: $q > 2$ and $1 < q < 2$.

The case of $q > 2$. For $q > 2$, we provide a quantum estimator with optimal sample complexity $\widetilde{\Theta}(1/\varepsilon^2)$ only up to a logarithmic factor in ε. This result is formally stated in the following theorem.

Theorem 1 (Optimal estimator for $q > 2$, Theorems 14 and 20) *For every $q > 2$, it is necessary and sufficient to use $\widetilde{\Theta}(1/\varepsilon^2)$ samples of the quantum state ρ to estimate $\mathrm{tr}(\rho^q)$ to within additive error ε.*

The case of $1 < q < 2$. For $1 < q < 2$, we provide a quantum estimator with sample complexity $\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$, only with room for quadratic improvements due to the lower bound $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1}, 2\}})$. This result is formally stated in the following theorem.

2. In Wang et al. (2024a), their main results only consider the quantum query complexity, as they assume access to the state-preparation circuit of ρ. Even though, their results also imply a sample complexity of $\mathrm{poly}(r, 1/\varepsilon)$ (with a polynomial overhead compared to the corresponding query complexity) using the techniques in Gilyén and Poremba (2022), as noted in (Wang et al., 2024a, Footnote 2).

Theorem 2 (Improved estimator for $1 < q < 2$, Theorems 16 and 20) *For every $1 < q < 2$, it is sufficient to use $\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$ samples of the quantum state ρ to estimate $\operatorname{tr}(\rho^q)$ to within additive error ε. On the other hand, when the dimension of ρ is sufficiently large, $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1},2\}})$ samples of ρ are necessary.*

Our estimators for Theorems 1 and 2 can be efficiently implemented with quantum time complexity $\operatorname{poly}(\log(d), 1/\varepsilon)$ for any constant $q > 1$ (see Section 3.4), where d is the dimension of ρ.

Both Theorems 1 and 2 improve the prior best upper bound $\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$ and lower bound $\Omega(1/\varepsilon)$ in Liu and Wang (2025). It is also noted that Theorem 1 gives a matching lower bound of $\Omega(1/\varepsilon^2)$ on the sample complexity of estimating $\operatorname{tr}(\rho^q)$ for every integer $q \geq 3$, implying that the Shift test in Ekert et al. (2002) is sample-optimal to estimate $\operatorname{tr}(\rho^q)$ to within an additive error, generalizing the lower bounds in Chen et al. (2023); Gong et al. (2024) for the optimality of the SWAP test Buhrman et al. (2001) to estimate $\operatorname{tr}(\rho^2)$. We summarize the developments for the sample complexity of estimating $\operatorname{tr}(\rho^q)$ in Table 1.

Table 1: Sample complexity of estimating $\operatorname{tr}(\rho^q)$.

$q \geq 2$	$1 < q < 2$	References
$O(1/\varepsilon^2)$, $q \in \mathbb{N}$	/	Buhrman et al. (2001); Ekert et al. (2002)
$\Omega(1/\varepsilon^2)$, $q = 2$	/	Chen et al. (2023); Gong et al. (2024)
$O(\operatorname{poly}(r, 1/\varepsilon))$		Acharya et al. (2020); Wang et al. (2024a,b) Wang and Zhang (2024a)
$\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$, $\Omega(1/\varepsilon)$		Liu and Wang (2025)
$\widetilde{\Theta}(1/\varepsilon^2)$	$\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$, $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1},2\}})$	This Work

1.2. Techniques

Upper bounds. Since the trace of quantum state power $\operatorname{tr}(\rho^q)$ is a unitarily invariant quantity, it is well-known that there exists a canonical estimator performing weak Schur sampling Childs et al. (2007); Montanaro and de Wolf (2016); O'Donnell and Wright (2021) on $\rho^{\otimes n}$ to obtain a Young diagram outcome λ and then predicting the final result $\operatorname{tr}(\rho^q)$ based on λ. The most straightforward way to do this is to treat each λ_i/n, where λ_i is the i-th row of λ, as an estimate of the i-th large eigenvalue of ρ, and then output $\sum_i(\lambda_i/n)^q$ as the final result, which is what is called the *plug-in estimator*. Existing quantum plug-in estimators are known for, e.g., von Neumann entropy and Rényi entropy in Acharya et al. (2020); Bavarian et al. (2016).

However, directly using the plug-in estimator with current error bounds for weak Schur sampling in O'Donnell and Wright (2017) seems to be difficult to avoid the dependence on the dimension (or rank) of ρ appearing in the accumulation of errors. This is very different from the classical empirical estimation. For example, the classical plug-in estimators for $\sum_i p_i^q$ in Jiao et al. (2015, 2017) suffice to achieve the optimal sample complexity, while the same strategy might introduce an *unexpected* factor of $\operatorname{poly}(d)$ in the quantum case, where d is the dimension. To overcome this limitation, we develop non-plug-in estimators for $\operatorname{tr}(\rho^q)$. Our non-plug-in estimator adopts a simple but effective truncation strategy which eliminates the dimension (or rank) in the complexity. Specifically, having obtained an estimated spectrum $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_d)$ of ρ to certain precision with $\hat{\alpha}_1 \geq \hat{\alpha}_2 \geq \cdots \geq \hat{\alpha}_d$ (with $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$ the true sorted spectrum of ρ), our non-plug-in estimator is then of the form

$$\hat{P} = \sum_{j=1}^{m} \hat{\alpha}_j^q,$$

where m is a truncation parameter such that the lower-order errors are controlled by the eigenvalues (which are finally suppressed due to constantly upper bounded partial sums), and the higher-order errors are accumulated with scaling only depending on m (thus suppressed with negligible truncation bias). In sharp contrast to the quantum plug-in estimators in the literature Acharya et al. (2020); Bavarian et al. (2016), our non-plug-in construction can be shown to achieve optimal sample complexity only up to a logarithmic factor (see Sections 3.2 and 3.3 for more details). As a result, we obtain sample upper bounds $\widetilde{O}(1/\varepsilon^2)$ for $q > 2$ and $\widetilde{O}(1/\varepsilon^{\frac{1}{q-1}})$ for $1 < q < 2$. Note that the exponent of the upper bound does not depend on q for constant $q > 2$, which is in contrast to $1 < q < 2$. This is because we borrow a factor α_i from α_i^q to control the error $|\hat{\alpha}_i - \alpha_i|$ (to avoid d-dependence), and the fluctuation of $\hat{\alpha}_i^{q-1}$ is small enough when $q > 2$ (see Equation (3)), causing q to disappear from the exponent.

Lower bounds. Our lower bounds consist of two parts: $\Omega(1/\varepsilon^{\frac{1}{q-1}})$ and $\Omega(1/\varepsilon^2)$.

The former lower bound $\Omega(1/\varepsilon^{\frac{1}{q-1}})$ for $1 < q < 2$ is obtained by reducing a discrimination task on ensembles of quantum states. Specifically, we consider two unitarily invariant ensembles of quantum states that are maximally mixed with respect to different dimensions. Then, we show that the discrimination between these ensembles can be characterized by the discrimination between certain Schur-Weyl distributions in their total variation distance. To bound the total variation distance, we recall the relationship between the Schur-Weyl distributions and Plancherel distributions shown in Childs et al. (2007), which demands a linear scaling with the dimensions. With carefully chosen dimension parameters, we can obtain our lower bound.

The latter lower bound $\Omega(1/\varepsilon^2)$ for any constant $q > 1$ is obtained by reducing from a state discrimination task with a simple but effective hard instance from Chen et al. (2023); Gong et al. (2024).

1.3. Related Work

After the work of Buhrman et al. (2001); Ekert et al. (2002), there have been a series of subsequent work focusing on the estimation of $\operatorname{tr}(\rho^q)$ for integer $q \geq 2$ Brun (2004); van Enk and Beenakker (2012); Johri et al. (2017); Subaşı et al. (2019); Yirka and Subaşı (2021); Quek et al. (2024); Zhou and Liu (2024); Shin et al. (2024). As the classical counterpart, estimating the functional $\sum_{j=1}^{N} p_j^q$ of a probability distribution p to within an additive error was studied in Antos and Kontoyiannis (2001) for integer $q \geq 2$, and later in Jiao et al. (2015, 2017) for non-integer q; its estimation to a multiplicative error was studied in Acharya et al. (2017) for Rényi entropy estimation. In addition, Shannon entropy estimation was studied in Paninski (2003, 2004); Valiant and Valiant (2011a,b, 2017); Wu and Yang (2016).

Given sample access to the quantum states to be tested, quantum estimators and testers for their properties have been investigated in the literature. The first optimal quantum tester was discovered in Childs et al. (2007), which distinguishes whether a quantum state has a spectrum uniform on

r or $2r$ eigenvalues. This was later generalized to an optimal tester for mixedness in O'Donnell and Wright (2021) and to quantum state certification in Bădescu et al. (2019). In addition, optimal estimators are known for Rényi entropy of integer order Acharya et al. (2020), and the closeness (trace distance and fidelity) between pure quantum states Wang and Zhang (2024b). A distributed optimal estimator was known for the inner product of quantum states Anshu et al. (2022). Estimators and testers with incoherent measurements are also known for purity Chen et al. (2021); Gong et al. (2024), unitarity Chen et al. (2021, 2023), certification Chen et al. (2022); Liu and Acharya (2024), and $\operatorname{tr}(\rho^q)$ for integer q (further used for spectrum estimation) Pelecanos et al. (2025). In addition to those that were known to be optimal, there are also estimators for entropy Acharya et al. (2020); Bavarian et al. (2016); Wang and Zhang (2024a); Liu and Wang (2025), relative entropy Hayashi (2025), fidelity Gilyén and Poremba (2022), and trace distance Wang and Zhang (2024c).

1.4. Discussion

In this paper, we presented quantum estimators for estimating $\operatorname{tr}(\rho^q)$ for non-integer $q > 1$, significantly improving the prior approaches. In particular, for $q > 2$, our estimators achieve optimal sample complexity only up to a logarithmic factor. Our (non-plug-in) estimators are directly constructed by weak Schur sampling with optimal sample complexity (although every estimator for unitarily invariant properties is known to imply a canonical estimator based on weak Schur sampling (Montanaro and de Wolf, 2016, Lemma 20)), in addition to the (plug-in) optimal estimator for Rényi entropy of integer order Acharya et al. (2020), the optimal testers for mixedness O'Donnell and Wright (2021) and quantum state certification Bădescu et al. (2019), and the optimal learners for full tomography Haah et al. (2017); O'Donnell and Wright (2016). At the end of the discussion, we list some questions in this direction for future research.

1. Can we remove the logarithmic factor from the sample complexity obtained in this paper?
2. Can we improve the upper or the lower bound for $1 < q < 2$?
3. Can we find more (plug-in or non-plug-in) optimal estimators based on weak Schur sampling?
4. Can we obtain optimal estimators for $\operatorname{tr}(\rho^q)$ with restricted measurements?
5. As the sample complexities of estimating $\operatorname{tr}(\rho^q)$ for $q \geq 2$ are known to be $\widetilde{\Theta}(1/\varepsilon^2)$ (thus they have almost the same difficulty in the sample complexity) but only the case of $q = 2$ is known to be BQP-hard Liu and Wang (2025), an interesting question is: can we show the BQP-hardness of estimating $\operatorname{tr}(\rho^q)$ for general $q > 2$?

2. Preliminaries

2.1. Basics in quantum computing

A d-dimensional (mixed) quantum state can be described by a $d \times d$ complex-valued positive semidefinite matrix $\rho \in \mathbb{C}^{d\times d}$ satisfying $\operatorname{tr}(\rho) = 1$. The trace distance between two quantum states ρ_0 and ρ_1 is defined by $\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \frac{1}{2}\operatorname{tr}(|\rho_0 - \rho_1|)$. The fidelity between two quantum states ρ_0 and ρ_1 is defined by $\mathrm{F}(\rho_0, \rho_1) = \operatorname{tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right)$. To discriminate two quantum states, we include the following well-known results. The following theorem can be found in (Wilde, 2013, Section 9.1.4), (Hayashi, 2016, Lemma 3.2), and (Watrous, 2018, Theorem 3.4).

Theorem 3 (Quantum state discrimination) *Any POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$ that distinguishes two quantum states ρ_0 and ρ_1 (each with a priori probability $1/2$) with success probability $\frac{1}{2}\operatorname{tr}(\Lambda_0\rho_0) + \frac{1}{2}\operatorname{tr}(\Lambda_1\rho_1) \le \frac{1}{2}\big(1 + \frac{1}{2}\|\rho_0 - \rho_1\|_1\big)$.*

The following fact was noted in (Haah et al., 2017, Section 1) using the quantum Chernoff bound Nussbaum and Szkoła (2009); Audenaert et al. (2007).

Fact 4 *The sample complexity for distinguishing two quantum states ρ_0 and ρ_1 is $\Omega(1/\gamma)$, where $\gamma = 1 - \mathrm{F}(\rho_0, \rho_1)$ is the infidelity.*

2.2. Basic representation theory

A *representation* of a group G is a pair $(\mu, \mathcal{H})$, where $\mathcal{H}$ is a (complex) Hilbert space, and $\mu : G \to \mathrm{GL}(\mathcal{H})$ is a group homomorphism from G to the general linear group on $\mathcal{H}$.[3] We also call $\mu(g)$ the action of $g \in G$ on $\mathcal{H}$. When the group action is clear from the context, we may omit μ and directly use $\mathcal{H}$ to refer to the representation of G.

A *sub-representation* of $(\mu, \mathcal{H})$ is a representation $(\mu', \mathcal{H}')$, where $\mathcal{H}'$ is a subspace of $\mathcal{H}$ and $\mu'(g)$ is the restriction of $\mu(g)$ to $\mathcal{H}'$. A representation $\mathcal{H}$ of G is *irreducible* if the only sub-representations of $\mathcal{H}$ are $\{0\}$ and $\mathcal{H}$ itself. A *representation homomorphism* between two representations $(\mu_1, \mathcal{H}_1), (\mu_2, \mathcal{H}_2)$ of group G is a linear operator $F : \mathcal{H}_1 \to \mathcal{H}_2$ which commutes with the action of G, i.e., $F\mu_1(g) = \mu_2(g)F$. A *representation isomorphism* is a representation homomorphism that is also a full-rank linear map. Two representations $\mathcal{H}_1$ and $\mathcal{H}_2$ of a group G are said to be *isomorphic* if there exists an representation isomorphism between them, and we write $\mathcal{H}_1 \overset{G}{\cong} \mathcal{H}_2$. Then, we introduce the Schur's Lemma, which is an important and basic result in representation theory.

Fact 5 (Schur's Lemma, see, e.g. (Etingof et al., 2011, Proposition 2.3.9)) *Let $\mathcal{H}_1, \mathcal{H}_2$ be irreducible representations of a group G. If $F : \mathcal{H}_1 \to \mathcal{H}_2$ is a non-zero homomorphism of representations, then F is an isomorphism.*

The following is a direct and useful corollary of Schur's Lemma.

Corollary 6 *Suppose $\mathcal{H}$ is an irreducible representation of G and $F : \mathcal{H} \to \mathcal{H}$ is a representation homomorphism. Then $F = cI$ where c is a complex number.*

2.2.1. SCHUR-WEYL DUALITY

A *Young diagram* λ with n boxes and at most d rows is a partition $\lambda = (\lambda_1, \ldots, \lambda_d)$ of n such that $\sum_i \lambda_i = n$ and $\lambda_1 \ge \cdots \ge \lambda_d \ge 0$. We use $\lambda \vdash n$ to denote that λ is a Young diagram with n boxes.

Consider the actions of the symmetric group $\mathfrak{S}_n$ and the unitary group $\mathbb{U}_d$ on the Hilbert space $(\mathbb{C}^d)^{\otimes n}$. For any $U \in \mathbb{U}_d$, U acts on $(\mathbb{C}^d)^{\otimes n}$ by $|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle \mapsto U|\psi_1\rangle \otimes \cdots \otimes U|\psi_n\rangle$, and for any $\pi \in \mathfrak{S}_n$, π acts on $(\mathbb{C}^d)^{\otimes n}$ by $|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle \mapsto |\psi_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |\psi_{\pi^{-1}(n)}\rangle$. For convenience, we directly use $U^{\otimes n}$ and π to denote the action of $U \in \mathbb{U}_d$ and $\pi \in \mathfrak{S}_n$ on $(\mathbb{C}^d)^{\otimes n}$.

Note that $U^{\otimes n}$ and π commutes with each others, which means $(\mathbb{C}^d)^{\otimes n}$ is also a representation of the group $\mathfrak{S}_n \times \mathbb{U}_d$. This is characterized by the following renowned Schur-Weyl duality.

3. In this paper, we mostly consider the case that G is finite or compact, where without loss of generality we can assume $\mu : G \to \mathbb{U}(\mathcal{H})$ is unitary.

Fact 7 (Schur-Weyl duality Fulton and Harris (2013); Etingof et al. (2011)) *$(\mathbb{C}^d)^{\otimes n}$ has the decomposition $(\mathbb{C}^d)^{\otimes n} \overset{\mathfrak{S}_n \times \mathbb{U}_d}{\cong} \bigoplus_{\lambda \vdash n} \mathcal{P}_\lambda \otimes \mathcal{Q}_\lambda^d$, where $\mathcal{P}_\lambda$ and $\mathcal{Q}_\lambda^d$ are irreducible representations of $\mathfrak{S}_n$ and $\mathbb{U}_d$, respectively, and are labeled by a Young diagram $\lambda \vdash n$.*[4]

For $\pi \in \mathfrak{S}_n$ and $U \in \mathbb{U}_d$, we use $\mathsf{p}_\lambda(\pi)$ and $\mathsf{q}_\lambda(U)$ to denote their actions on $\mathcal{P}_\lambda$ and $\mathcal{Q}_\lambda^d$, respectively.

Remark 8 *In fact, q_λ can be extended naturally to the actions of the group $\mathrm{GL}(\mathbb{C}^d)$ on $\mathcal{Q}_\lambda^d$, and further by continuity to the action of any matrix in $\mathrm{End}(\mathbb{C}^d)$ on $\mathcal{Q}_\lambda^d$.*

For any matrix $X \in \mathrm{End}(\mathbb{C}^d)$, $X^{\otimes n}$ is invariant under permutations (the actions of $\mathfrak{S}_n$). It is not hard using Schur's Lemma to show the following fact.

Fact 9 *$X^{\otimes n}$ has the following form: $X^{\otimes n} = \bigoplus_{\lambda \vdash n} I_{\mathcal{P}_\lambda} \otimes \mathsf{q}_\lambda(X)$, where $\mathsf{q}_\lambda(X)$ is the action of X on $\mathcal{Q}_\lambda^d$ (see Remark 8).*

Furthermore, it is known that $\mathrm{tr}(\mathsf{q}_\lambda(X)) = s_\lambda(\alpha)$, where s_λ is the *Schur polynomial* Fulton and Harris (2013) indexed by λ and $\alpha = (\alpha_1, \ldots, \alpha_d)$ are the eigenvalues of X.

2.3. Weak Schur sampling as quantum estimators

Suppose we have n samples of an unknown d-dimensional quantum state ρ. Consider the task of estimating a quantitative property $F(\rho)$ of ρ (e.g., the purity $\mathrm{tr}(\rho^2)$). Without loss of generality, the estimator can be described by a POVM $\{M_i\}$ applied on $\rho^{\otimes n}$,[5] and $f(i)$ is returned as an estimate if the measurement outcome is i.

Note that $\rho^{\otimes n}$ is invariant under permutations of the tensors, i.e., for any $\pi \in \mathfrak{S}_n$, $\pi\rho^{\otimes n}\pi^\dagger = \rho^{\otimes n}$. This means we can "factor out" the action of the symmetric group $\mathfrak{S}_n$ to obtain a permutation invariant estimator. Furthermore, if the quantitative property $F(\rho)$ is unitarily invariant, i.e., for any $U \in \mathbb{U}_d$, $F(U\rho U^\dagger) = F(\rho)$, we can also factor out the action of the unitary group $\mathbb{U}_d$ to obtain a unitarily invariant estimator with the performance no worse than the original one. Specifically, we define the canonical permutation-invariant and unitary-invariant estimator $\{\overline{M}_i\}$ as: $\overline{M}_i = \frac{1}{n!}\sum_{\pi \in \mathfrak{S}_n} \pi\, \mathbb{E}_{U \in \mathbb{U}_d}\left[U^{\otimes n} M_i U^{\dagger \otimes n}\right]\pi^\dagger$. The following shows that the estimator $\{\overline{M}_i\}_i$ is at least as powerful as the original estimator $\{M_i\}_i$ (see also, e.g., Montanaro and de Wolf (2016); Hayashi (2025)).

Fact 10 *If $\{M_i\}$ is an estimator of the quantitative property F achieving additive error ε with success probability $1-\delta$, then $\{\overline{M}_i\}$ can also achieve additive error ε with probability $1-\delta$.*

Note that $\overline{M}_i$ commutes with both $U^{\otimes n}$ and π for any $U \in \mathbb{U}_d$ and $\pi \in \mathfrak{S}_n$. By the Schur-Weyl duality (see Fact 7) and Corollary 6, we have $\overline{M}_i = \bigoplus_{\lambda \vdash n} c_{i,\lambda} \cdot I_{\mathcal{P}_\lambda} \otimes I_{\mathcal{Q}_\lambda^d}$, where $c_{i,\lambda}$ is a positive number such that $\sum_i c_{i,\lambda} = 1$. Then, by Fact 9, we can see that the estimator $\{\overline{M}_i\}$ applied on $\rho^{\otimes n}$ is equivalent to

1. sample a $\lambda \vdash n$ from the distribution $\{\mathrm{tr}(I_{\mathcal{P}_\lambda} \otimes \mathsf{q}_\lambda(\rho))\}_\lambda = \{\dim(\mathcal{P}_\lambda) \cdot s_\lambda(\alpha)\}_\lambda$, where s_λ is the Schur polynomial and $\alpha = (\alpha_1, \ldots, \alpha_d)$ are the eigenvalues of ρ such that $\alpha_1 \geq \cdots \geq \alpha_d$.

4. Note that if the Young diagram λ has more than d rows, then $\mathcal{Q}_\lambda^d = 0$.
5. Here, we assume the POVM is discrete, the continuous case can be treated similarly.

2. sample an i from the distribution $\{c_{i,\lambda}\}_i$.

It is worth noting that, the second step is entirely classical, while the first step is a quantum measurement independent of the specific task, which is called *weak Schur sampling* Childs et al. (2007). In step 1, the distribution $\{\dim(\mathcal{P}_\lambda) \cdot s_\lambda(\alpha)\}_\lambda$ is referred to as the *Schur-Weyl distribution* O'Donnell and Wright (2017) and is denoted by $\mathrm{SW}^n(\alpha)$ or $\mathrm{SW}^n(\rho)$. Specifically,$\Pr_{\lambda' \sim \mathrm{SW}^n(\alpha)}[\lambda' = \lambda] = \dim(\mathcal{P}_\lambda) \cdot s_\lambda(\alpha)$. Furthermore, the Young diagram $\lambda \sim \mathrm{SW}^n(\alpha)$ provides a good approximation of the eigenvalues $\alpha_1, \ldots, \alpha_d$ of ρ, which is characterized by the following result.

Lemma 11 (Adapted from (O'Donnell and Wright, 2017, Theorem 1.5)) *For $j \in [d]$, we have $\mathbb{E}_{\lambda \sim \mathrm{SW}^n(\alpha)}\left[(\lambda_j - \alpha_j n)^2\right] \le O(n)$.*

We use SW^n_d to denote $\mathrm{SW}^n(\alpha)$ when $\alpha = (1/d, \ldots, 1/d)$,[6] i.e., ρ is maximally mixed. Furthermore, when $d \to \infty$, the distribution tends to a limiting distribution $\mathrm{Planch}(n)$, called *Plancherel distribution* over the symmetric group $\mathfrak{S}_n$. We will use the following result which provides both upper and lower bounds of the convergence of SW^n_d to $\mathrm{Planch}(n)$.

Lemma 12 ((Childs et al., 2007, Lemma 6)) *If $2 \le n \le d$, then $\frac{n}{36d} \le \| \mathrm{SW}^n_d - \mathrm{Planch}(n)\|_1 \le \sqrt{2}\frac{n}{d}$.*

3. Upper Bounds

In this section, we provide quantum algorithms that estimate the value of $\mathrm{tr}(\rho^q)$ for $q > 2$ and $1 < q < 2$ respectively in Section 3.2 and Section 3.3. For these, we also provide a simple approach to the quantum spectrum estimation with entry-wise bounds in Section 3.1.

In both of our quantum algorithms, we use the following three parameters m, δ', ε', where m is the position where the truncation is taken, and δ' and ε' are, respectively, the failure probability and the precision when applying the quantum spectrum estimation with entry-wise bounds in Section 3.1. Specifically, $m \in [d]$ is a positive integer and $\delta', \varepsilon' \in (0, 1)$ are real numbers, all of which are to be determined later. In addition, we assume that ρ has the spectrum decomposition:$\rho = \sum_{j=1}^d \alpha_j |\psi_j\rangle\langle\psi_j|$, where $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_d \ge 0$ with $\sum_{j=1}^d \alpha_j = 1$ and $\{|\psi_j\rangle\}$ is an orthonormal basis.

3.1. Quantum spectrum estimation with entry-wise bounds

Efficient approaches to quantum spectrum estimation were given in O'Donnell and Wright (2016) in the ℓ_1 and ℓ_2 distances and in O'Donnell and Wright (2017) in the Hellinger-squared distance, chi-squared divergence, and Kullback-Liebler (KL) divergence. In this section, we provide an efficient approach to quantum spectrum estimation with entry-wise bounds based on the results of O'Donnell and Wright (2017), which will be used as a subroutine in our estimators for $\mathrm{tr}(\rho^q)$ in Section 3.

Lemma 13 (Quantum spectrum estimation with entry-wise bounds) *For every $\varepsilon, \delta \in (0, 1)$, we can use $O(\log(1/\delta)/\varepsilon^2)$ samples of ρ to obtain a sequence of random variables $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d) \in \mathbb{R}^d$ such that for every $j \in [d]$, it holds with probability at least $1 - \delta$ that $|\hat{\alpha}_j - \alpha_j| \le \varepsilon$.*

Proof The formal algorithm is given in Algorithm 1. The full proof is given in Appendix A. ■

6. In some papers, SW^n_d is also called Schur-Weyl distribution O'Donnell and Wright (2021) or simply Schur distribution Childs et al. (2007).

Algorithm 1 `SpectrumEstimation`$(\rho, \varepsilon, \delta)$

Input: Sample access to a d-dimensional mixed quantum state ρ; $\varepsilon, \delta \in (0, 1)$.
Output: A d-dimensional vector $\hat{\alpha} \in \mathbb{R}^d$.

$n \leftarrow \Theta(1/\varepsilon^2)$, $k \leftarrow \Theta(\log(1/\delta))$.
for $l = 1, 2, \ldots, k$
  $\lambda^{(l)} \sim \mathrm{SW}^n(\rho)$.
end
for $j = 1, 2, \ldots, d$
  $\hat{\alpha}_j \leftarrow \mathrm{median}\{\underline{\lambda}_j^{(1)}, \underline{\lambda}_j^{(2)}, \ldots, \underline{\lambda}_j^{(k)}\}$, where $\underline{\lambda}_j^{(l)} = \lambda_j^{(l)}/n$.
end
return $(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d)$.

3.2. $q > 2$

Algorithm 2 `PowerTrace`(ρ, q, ε) for $q > 2$

Input: Sample access to a d-dimensional mixed quantum state ρ; $q \in (2, +\infty)$ and $\varepsilon \in (0, 1)$.
Output: An estimate of $\mathrm{tr}(\rho^q)$.

$\varepsilon' \leftarrow \varepsilon/(q+3)$, $m \leftarrow \min\{\lceil 1/\varepsilon' \rceil, d\}$, $\delta' \leftarrow 1/3m$.
$\hat{\alpha} \leftarrow$ `SpectrumEstimation`$(\rho, \varepsilon', \delta')$.
$\hat{P} \leftarrow \sum_{j=1}^{m} \hat{\alpha}_j^q$.
return $\hat{P}$.

For $q > 2$, the sample complexity of estimating $\mathrm{tr}(\rho^q)$ is given as follows.

Theorem 14 *For every constant $q > 2$, we can estimate $\mathrm{tr}(\rho^q)$ to within additive error ε using $O(\log(1/\varepsilon)/\varepsilon^2)$ samples of ρ.*

Our estimator for Theorem 14 is formally given in Algorithm 2. To prove Theorem 14, we need the following inequalities.

Fact 15 *For $\alpha > 1$ and $x, y \in [0, 1]$, we have $x^\alpha \leq x$ and $|x^\alpha - y^\alpha| \leq \alpha|x - y|$.*

Proof This fact follows by applying the mean value theorem on the function $f(x) = x^\alpha$. ■

Now we are ready to prove Theorem 14.

Proof [Proof of Theorem 14] Let parameters $m \in \mathbb{N}$ and $\delta', \varepsilon' \in (0, 1)$ to be determined later. By Lemma 13, we can use $O(\log(1/\delta')/\varepsilon'^2)$ samples of ρ to obtain a sequence $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d)$ such that for every $j \in [d]$,

$$\Pr\left[|\hat{\alpha}_j - \alpha_j| \leq \varepsilon'\right] \geq 1 - \delta'. \tag{1}$$

Then, we consider the estimator $\hat{P} \coloneqq \sum_{j=1}^{m} \hat{\alpha}_j^q$. The additive error is bounded by

$$\left|\hat{P} - \mathrm{tr}(\rho^q)\right| = \left|\sum_{j=1}^{m}\left(\hat{\alpha}_j^q - \alpha_j^q\right) - \sum_{j=m+1}^{d} \alpha_j^q\right| \leq \sum_{j=1}^{m}\left|\hat{\alpha}_j^q - \alpha_j^q\right| + \sum_{j=m+1}^{d} \alpha_j^q. \tag{2}$$

For the first term of Equation (2), note that $\hat{\alpha}_j^q - \alpha_j^q = (\hat{\alpha}_j - \alpha_j)\hat{\alpha}_j^{q-1} + \alpha_j\hat{\alpha}_j^{q-1} - \alpha_j^q = (\hat{\alpha}_j - \alpha_j)\hat{\alpha}_j^{q-1} + \alpha_j(\hat{\alpha}_j^{q-1} - \alpha_j^{q-1})$, then we have

$$|\hat{\alpha}_j^q - \alpha_j^q| \le |\hat{\alpha}_j - \alpha_j||\hat{\alpha}_j|^{q-1} + |\alpha_j||\hat{\alpha}_j^{q-1} - \alpha_j^{q-1}| \le |\hat{\alpha}_j - \alpha_j||\hat{\alpha}_j| + |\alpha_j|(q-1)|\hat{\alpha}_j - \alpha_j| \quad (3)$$
$$\le q\alpha_j|\hat{\alpha}_j - \alpha_j| + |\hat{\alpha}_j - \alpha_j|^2, \quad (4)$$

where Equation (3) is by Fact 15. From Equation (4) and by Equation (1), the following holds with probability $\ge 1 - \delta'$: $|\hat{\alpha}_j^q - \alpha_j^q| \le q\alpha_j\varepsilon' + \varepsilon'^2$. Therefore, we have that with probability $\ge 1 - m\delta'$, the following holds:

$$\sum_{j=1}^{m}\left|\hat{\alpha}_j^q - \alpha_j^q\right| \le \sum_{j=1}^{m}\left(q\alpha_j\varepsilon' + \varepsilon'^2\right) = q\varepsilon'\sum_{j=1}^{m}\alpha_j + m\varepsilon'^2 \le q\varepsilon' + m\varepsilon'^2. \quad (5)$$

On the other hand, by noting that $\alpha_j \le 1/j$ (since $j\alpha_j \le \alpha_1 + \cdots + \alpha_j \le 1$) for every $j \in [d]$, we have

$$\sum_{j=m+1}^{d}\alpha_j^q \le \sum_{j=m+1}^{d}\left(\frac{1}{j}\right)^q \le \int_m^d \left(\frac{1}{x}\right)^q \mathrm{d}x = \frac{m^{1-q} - d^{1-q}}{q-1}. \quad (6)$$

Combining Equations (5) and (6) in Equation (2), we have that with probability $\ge 1 - m\delta'$, the following holds:

$$\left|\hat{P} - \operatorname{tr}(\rho^q)\right| \le q\varepsilon' + m\varepsilon'^2 + \frac{m^{1-q} - d^{1-q}}{q-1}. \quad (7)$$

By taking $\varepsilon' := \frac{\varepsilon}{q+3}, m = \min\left\{\left\lceil\frac{1}{\varepsilon'}\right\rceil, d\right\}, \delta' := \frac{1}{3m}$, we have from Equation (7) that with probability $\ge 1 - m\delta' = 2/3$, it holds that $\left|\hat{P} - \operatorname{tr}(\rho^q)\right| \le \varepsilon$. To see this, we consider the following two cases:

1. $1/\varepsilon' \le d$. In this case, $1/\varepsilon' \le m = \lceil 1/\varepsilon' \rceil < 1/\varepsilon' + 1$. We have (7) $\le q\varepsilon' + (1/\varepsilon' + 1)\varepsilon'^2 + 1/m \le q\varepsilon' + \varepsilon' + \varepsilon'^2 + \varepsilon' \le (q+3)\varepsilon' = \varepsilon$.
2. $1/\varepsilon' > d$. In this case, $m = d < 1/\varepsilon'$. We have (7) $= q\varepsilon' + d\varepsilon'^2 \le (q+1)\varepsilon' < \varepsilon$.

To complete the proof, the sample complexity is $O(\log(1/\delta')/\varepsilon'^2) = O(\log(1/\varepsilon)/\varepsilon^2)$. ■

3.3. $1 < q < 2$

We state the sample complexity of estimating $\operatorname{tr}(\rho^q)$ for the case of $1 < q < 2$ as follows.

Theorem 16 *For every constant $1 < q < 2$, we can estimate $\operatorname{tr}(\rho^q)$ to within additive error ε using $O(\log(1/\varepsilon)/\varepsilon^{\frac{2}{q-1}})$ samples of ρ.*

Our estimator for Theorem 16 is formally given in Algorithm 3. To show Theorem 16, we need the following inequalities.

Algorithm 3 $\texttt{PowerTrace}(\rho, q, \varepsilon)$ for $1 < q < 2$

Input: Sample access to a d-dimensional mixed quantum state ρ; $q \in (1,2)$ and $\varepsilon \in (0,1)$.
Output: An estimate of $\mathrm{tr}(\rho^q)$.

1: $\varepsilon' \leftarrow (\varepsilon/5)^{\frac{1}{q-1}}$, $m \leftarrow \min\{\lceil 1/\varepsilon' \rceil, d\}$, $\delta' \leftarrow 1/3m$.
2: $\hat{\alpha} \leftarrow \texttt{SpectrumEstimation}(\rho, \varepsilon', \delta')$.
3: $\hat{P} \leftarrow \sum_{j=1}^{m} \hat{\alpha}_j^q$.
4: **return** $\hat{P}$.

Fact 17 *For $0 \le x \le y \le 1$ and $0 < s < 1$ we have $y^s - x^s \le (y-x)^s$.*

Proof This fact follows by considering the derivative of the function $f(x) \coloneqq (y-x)^s + x^s$. ■

Fact 18 (By Roger–Hölder's inequality Roger (1888); Hölder (1889)) *For $0 < s < 1$ and $x_i \ge 0$, we have $\sum_{i=1}^{k} x_i^s \le k^{1-s} \cdot (\sum_{i=1}^{k} x_i)^s$.*

Lemma 19 *Suppose that $1 < q < 2$ and $x_1 \ge x_2 \ge \cdots \ge x_N \ge 0$ with $\sum_{i=1}^{N} x_i = 1$. For any positive integer $m \le N$, we have $\sum_{i=m+1}^{N} x_i^q \le \frac{1}{m^{q-1}}$.*

Proof The proof can be found in Appendix B. ■

Now we are ready to prove Theorem 16.

Proof [Proof of Theorem 16] Let parameters $m \in \mathbb{N}$ and $\delta', \varepsilon' \in (0,1)$ to be determined later. By Lemma 13, we can use $O(\log(1/\delta')/\varepsilon'^2)$ samples of ρ to obtain a sequence $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_d)$ such that for every $j \in [d]$,

$$\Pr\left[|\hat{\alpha}_j - \alpha_j| \le \varepsilon'\right] \ge 1 - \delta'. \tag{8}$$

Then, we consider the estimator: $\hat{P} \coloneqq \sum_{j=1}^{m} \hat{\alpha}_j^q$. We have

$$\left|\hat{P} - \mathrm{tr}(\rho^q)\right| = \left|\sum_{j=1}^{m} \left(\hat{\alpha}_j^q - \alpha_j^q\right) - \sum_{j=m+1}^{d} \alpha_j^q\right| \le \sum_{j=1}^{m} \left|\hat{\alpha}_j^q - \alpha_j^q\right| + \sum_{j=m+1}^{d} \alpha_j^q. \tag{9}$$

For the first term of Equation (9), note that

$$\begin{aligned} \left|\hat{\alpha}_j^q - \alpha_j^q\right| &= \left|(\hat{\alpha}_j - \alpha_j)\hat{\alpha}_j^{q-1} + \alpha_j(\hat{\alpha}_j^{q-1} - \alpha_j^{q-1})\right| \le |\hat{\alpha}_j - \alpha_j|\hat{\alpha}_j^{q-1} + \alpha_j\left|\hat{\alpha}_j^{q-1} - \alpha_j^{q-1}\right| \\ &\le |\hat{\alpha}_j - \alpha_j|\hat{\alpha}_j^{q-1} + \alpha_j|\hat{\alpha}_j - \alpha_j|^{q-1}, \end{aligned} \tag{10}$$

where the last inequality is by Fact 17. Then, by Equation (8), with probability $\ge 1 - \delta'$, the following holds: (10) $\le \varepsilon'(\alpha_j + \varepsilon')^{q-1} + \alpha_j(\varepsilon')^{q-1}$. This implies, with probability $\ge 1 - m\delta'$, we have

$$\begin{aligned} \sum_{j=1}^{m}\left|\hat{\alpha}_j^q - \alpha_j^q\right| &\le \varepsilon' \sum_{j=1}^{m}(\alpha_j + \varepsilon')^{q-1} + (\varepsilon')^{q-1}\sum_{j=1}^{m}\alpha_j \le \varepsilon' \sum_{j=1}^{m}(\alpha_j + \varepsilon')^{q-1} + (\varepsilon')^{q-1} \\ &\le \varepsilon' m^{2-q} \cdot \left(m\varepsilon' + \sum_{j=1}^{m}\alpha_j\right)^{q-1} + (\varepsilon')^{q-1} \qquad (11) \\ &\le \varepsilon' m^{2-q}\left(m\varepsilon' + 1\right)^{q-1} + (\varepsilon')^{q-1}, \qquad (12) \end{aligned}$$

where Equation (11) is by Fact 18.

Combining Equation (12) with Equation (9), we have that, with probability $\geq 1 - m\delta'$, it holds that

$$\left|\hat{P} - \operatorname{tr}(\rho^q)\right| \leq \varepsilon' m^{2-q}\big(m\varepsilon' + 1\big)^{q-1} + (\varepsilon')^{q-1} + \sum_{j=m+1}^{d} \alpha_j^q. \tag{13}$$

By taking $\varepsilon' := \left(\frac{\varepsilon}{5}\right)^{\frac{1}{q-1}}$, $m = \min\left\{\left\lceil \frac{1}{\varepsilon'} \right\rceil, d\right\}$, $\delta' := \frac{1}{3m}$, we have from Equation (13) that with probability $\geq 1 - m\delta' = 2/3$, it holds that $\left|\hat{P} - \operatorname{tr}(\rho^q)\right| \leq \varepsilon$. To see this, we consider the following two cases:

1. $1/\varepsilon' \leq d$. In this case, $1/\varepsilon' \leq m = \lceil 1/\varepsilon' \rceil < 2/\varepsilon'$. We use Lemma 19 to obtain: $\sum_{j=m+1}^{d} \alpha_j^q \leq \frac{1}{m^{q-1}}$. Then, we have (13) $\leq \varepsilon'(2/\varepsilon')^{2-q}(2+1)^{q-1} + (\varepsilon')^{q-1} + (1/\varepsilon')^{1-q} \leq 3(\varepsilon')^{q-1} + 2(\varepsilon')^{q-1} \leq 5(\varepsilon')^{q-1} = \varepsilon$.
2. $1/\varepsilon' > d$. In this case, $m = d < 1/\varepsilon'$. We have (13) $\leq \varepsilon'(1/\varepsilon')^{2-q}(1+1)^{q-1} + (\varepsilon')^{q-1} \leq 5(\varepsilon')^{q-1} = \varepsilon$.

To complete the proof, the sample complexity is $O(\log(1/\delta')/\varepsilon'^2) = O(\log(1/\varepsilon)/\varepsilon^{\frac{2}{q-1}})$. ■

3.4. Time efficiency

Our estimators in Theorems 14 and 16 can actually be implemented with quantum time complexity $\operatorname{poly}(\log(d), 1/\varepsilon)$ for any constant $q > 1$. This is because, in Algorithms 2 and 3, we only need the first m entries of the output of Algorithm 1, where $m \leq O(1/\varepsilon^{\max\{1, \frac{1}{q-1}\}})$. On the other hand, Algorithm 1 uses $n = \widetilde{O}(1/\varepsilon^{\max\{\frac{2}{q-1}, 2\}})$ samples of ρ and can be implemented with quantum time complexity $O(n^3 \operatorname{polylog}(n, d)) = \widetilde{O}(1/\varepsilon^{\max\{\frac{6}{q-1}, 6\}}) \cdot \operatorname{polylog}(d)$ by weak Schur sampling.[7]

4. Lower Bounds

In this section, we prove a lower bound of $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1}, 2\}})$ on the sample complexity of estimating $\operatorname{tr}(\rho^q)$ for $q > 1$.

Theorem 20 *For any constant $q > 1$, any quantum estimator to additive error ε for $\operatorname{tr}(\rho^q)$ requires sample complexity $\Omega(1/\varepsilon^2)$. Moreover, for $1 < q < 2$, when the dimension of ρ is sufficiently large, it requires sample complexity $\Omega(1/\varepsilon^{\frac{1}{q-1}})$.*

Proof The proof of $\Omega(1/\varepsilon^2)$ is in Appendix C. Here, we only prove the lower bound $\Omega(1/\varepsilon^{\frac{1}{q-1}})$ for $1 < q < 2$. For integers $1 \leq r \leq d$, let $D_{r,d}$ denote the $d \times d$ diagonal matrix: $D_{r,d} := \operatorname{diag}(\underbrace{\tfrac{1}{r}, \ldots, \tfrac{1}{r}}_{r}, \underbrace{0, \ldots, 0}_{d-r})$. Let $r = \lfloor 1/(2\varepsilon)^{\frac{1}{q-1}} \rfloor$ and $d = \lfloor 1/\varepsilon^{\frac{1}{q-1}} \rfloor + 1$. If the number of samples $n > r$, then we directly have $n \geq \Omega(1/\varepsilon^{\frac{1}{q-1}})$. Therefore, we assume $n \leq r$. Then, consider the following problem.

7. This quantum time complexity was noted in Wang and Zhang (2024c,a); Hayashi (2025). This is achieved by using the implementation of weak Schur sampling introduced in (Montanaro and de Wolf, 2016, Section 4.2.2), equipped with the quantum Fourier transform over symmetric groups in Kawano and Sekigawa (2016).

Problem 1 *Suppose a d-dimensional quantum state ρ is in one of the following with equal probability:* ***1)*** $\rho = \rho_1 \coloneqq UD_{r,d}U^\dagger$*, where* $U \sim \mathbb{U}_d$ *is a d-dimensional Haar random unitary;* ***2)*** $\rho = \rho_2 \coloneqq D_{d,d}$*. The task is to distinguish between the above two cases.*

Note that $\operatorname{tr}(\rho_1^q) = 1/r^{q-1} \geq 2\varepsilon$ and $\operatorname{tr}(\rho_2^q) = 1/d^{q-1} \leq \varepsilon$. Therefore, any estimator of $\operatorname{tr}(\rho^q)$ to additive error $\frac{1}{2}\varepsilon = \Theta(\varepsilon)$ is able to distinguish the two cases in Problem 1.

On the other hand, suppose we have n samples of ρ. Then, for the first case (i.e., $\rho = \rho_1$), we have

$$\mathbb{E}\left[\rho_1^{\otimes n}\right] = \mathop{\mathbb{E}}_{U\sim\mathbb{U}_d}\left[U^{\otimes n}D_{r,d}^{\otimes n}U^{\dagger\otimes n}\right] = \sum_{\lambda\vdash n} I_{\mathcal{P}_\lambda} \otimes \mathop{\mathbb{E}}_{U\sim\mathbb{U}_d}\left[\mathsf{q}_\lambda(U)\mathsf{q}_\lambda(D_{r,d})\mathsf{q}_\lambda(U)^\dagger\right] \tag{14}$$

$$= \sum_{\lambda\vdash n} I_{\mathcal{P}_\lambda} \otimes I_{\mathcal{Q}_\lambda^d} \cdot \frac{s_\lambda(D_{r,d})}{\dim(\mathcal{Q}_\lambda^d)}, \tag{15}$$

where Equation (14) can be seen by Fact 9, in Equation (15) is by Corollary 6 and the observation that $\mathbb{E}_{U\sim\mathbb{U}_d}\left[\mathsf{q}_\lambda(U)\mathsf{q}_\lambda(D_{r,d})\mathsf{q}_\lambda(U)^\dagger\right]$ commutes with the actions of $U \in \mathbb{U}_d$, in which $s_\lambda(D_{r,d})$ refers to $s_\lambda(\underbrace{1/r,\ldots,1/r}_{r},\underbrace{0,\ldots,0}_{d-r})$. Similarly, for the second case (i.e., $\rho = \rho_2$), we have $\mathbb{E}\left[\rho_2^{\otimes n}\right] = \sum_{\lambda\vdash n} I_{\mathcal{P}_\lambda} \otimes I_{\mathcal{Q}_\lambda^d} \cdot \frac{s_\lambda(D_{d,d})}{\dim(\mathcal{Q}_\lambda^d)}$. By Theorem 3, the success probability of distinguishing $\mathbb{E}[\rho_1^{\otimes n}]$ and $\mathbb{E}[\rho_2^{\otimes n}]$ is upper bounded by $\frac{1}{2} + \frac{1}{4}\left\|\mathbb{E}\left[\rho_1^{\otimes n}\right] - \mathbb{E}\left[\rho_2^{\otimes n}\right]\right\|_1$. Note that

$$\left\|\mathbb{E}\left[\rho_1^{\otimes n}\right] - \mathbb{E}\left[\rho_2^{\otimes n}\right]\right\|_1 = \sum_{\lambda\vdash n} \left|\dim(\mathcal{P}_\lambda)s_\lambda(D_{r,d}) - \dim(\mathcal{P}_\lambda)s_\lambda(D_{d,d})\right| = \|\mathrm{SW}_r^n - \mathrm{SW}_d^n\|_1, \tag{16}$$

where in Equation (16) we use the stability of Schur polynomial, i.e., $s_\lambda(D_{r,d}) = s_\lambda(\underbrace{1/r,\ldots,1/r}_{r}, \underbrace{0,\ldots,0}_{d-r}) = s_\lambda(\underbrace{1/r,\ldots,1/r}_{r})$. Then, since $n \leq r \leq d$, by Lemma 12, we have that

$$\frac{n}{36r} \leq \|\,\mathrm{SW}_r^n - \mathrm{Planch}(n)\|_1 \leq \sqrt{2}\frac{n}{r}, \quad \text{and} \quad \frac{n}{36d} \leq \|\,\mathrm{SW}_d^n - \mathrm{Planch}(n)\|_1 \leq \sqrt{2}\frac{n}{d}.$$

This means $\|\,\mathrm{SW}_r^n - \mathrm{SW}_d^n\,\|_1 \leq \|\,\mathrm{SW}_r^n - \mathrm{Planch}(n)\|_1 + \|\,\mathrm{SW}_d^n - \mathrm{Planch}(n)\|_1 \leq \sqrt{2}\frac{n}{r} + \sqrt{2}\frac{n}{d}$.

Therefore, if the success probability is at least $2/3$, then $\frac{2}{3} \leq \frac{1}{2} + \frac{1}{4}\left(\sqrt{2}\frac{n}{r} + \sqrt{2}\frac{n}{d}\right) \leq \frac{1}{2} + \frac{n}{\sqrt{2}r}$, which means $n \geq \Omega(r) = \Omega(1/\varepsilon^{\frac{1}{q-1}})$. ∎

Acknowledgment

The authors would like to thank Yupan Liu for valuable comments and suggesting Question 5 in the Discussion.

The work of Qisheng Wang was supported by the Engineering and Physical Sciences Research Council under Grant EP/X026167/1.

References

Jayadev Acharya, Alon Orlitsky, Ananda Theertha Suresh, and Himanshu Tyagi. Estimating Renyi entropy of discrete distributions. *IEEE Transactions on Information Theory*, 63(1):38–56, 2017. doi: 10.1109/TIT.2016.2620435.

Jayadev Acharya, Ibrahim Issa, Nirmal V. Shende, and Aaron B. Wagner. Estimating quantum entropy. *IEEE Journal on Selected Areas in Information Theory*, 1(2):454–468, 2020. doi: 10.1109/JSAIT.2020.3015235.

Anurag Anshu, Zeph Landau, and Yunchao Liu. Distributed quantum inner product estimation. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 44–51, 2022. doi: 10.1145/3519935.3519974.

András Antos and Ioannis Kontoyiannis. Convergence properties of functional estimates for discrete distributions. *Random Structures & Algorithms*, 19(3–4):163–193, 2001. doi: 10.1002/rsa.10019.

K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Physical Review Letters*, 98(16): 160501, 2007. doi: 10.1103/PhysRevLett.98.160501.

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

Mohammad Bavarian, Saeed Mehraban, and John Wright. Learning entropy. A manuscript on von Neumann entropy estimation, private communication, 2016.

Christian Beck. Dynamical foundations of nonextensive statistical mechanics. *Physical Review Letters*, 87(18):180601, 2001. doi: 10.1103/PhysRevLett.87.180601.

Christian Beck. Generalized statistical mechanics and fully developed turbulence. *Physica A: Statistical Mechanics and its Applications*, 306:189–198, 2002. doi: 10.1016/S0378-4371(02)00497-1.

Todd A. Brun. Measuring polynomial functions of states. *Quantum Information and Computation*, 4(5):401–408, 2004. doi: 10.26421/QIC4.5-6.

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.

Kean Chen, Qisheng Wang, Peixun Long, and Mingsheng Ying. Unitarity estimation for quantum channels. *IEEE Transactions on Information Theory*, 69(8):5116–5134, 2023. doi: 10.1109/TIT.2023.3263645.

Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *Proceedings of the 62nd IEEE Annual Symposium on Foundations of Computer Science*, pages 574–585, 2021. doi: 10.1109/FOCS52979.2021.00063.

Sitan Chen, Brice Huang, Jerry Li, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. In *Proceedings of the 63rd IEEE Annual Symposium on Foundations of Computer Science*, pages 1205–1213, 2022. doi: 10.1109/FOCS54457.2022.00118.

Andrew M. Childs, Aram W. Harrow, and Pawel Wocjan. Weak Fourier-Schur sampling, the hidden subgroup problem, and the quantum collision problem. In *Proceedings of the 24th Annual Symposium on Theoretical Aspects of Computer Science*, pages 598–609, 2007. doi: 10.1007/978-3-540-70918-3_51.

Artur K. Ekert, Carolina Moura Alves, Daniel K. L. Oi, Michał Horodecki, Paweł Horodecki, and L. C. Kwek. Direct estimations of linear and nonlinear functionals of a quantum state. *Physical Review Letters*, 88(21):217901, 2002. doi: 10.1103/PhysRevLett.88.217901.

Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina. *Introduction to Representation Theory*, volume 59 of *Student Mathematical Library*. American Mathematical Society, 2011. doi: 10.1090/stml/059.

William Fulton and Joe Harris. *Representation Theory: A First Course*, volume 129 of *Graduate Texts in Mathematics*. Springer, 2013. doi: 10.1007/978-1-4612-0979-9.

András Gilyén and Alexander Poremba. Improved quantum algorithms for fidelity estimation. ArXiv e-prints, 2022.

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 Annual ACM SIGACT Symposium on Theory of Computing*, pages 193–204, 2019. doi: 10.1145/3313276.3316366.

Weiyuan Gong, Jonas Haferkamp, Qi Ye, and Zhihan Zhang. On the sample complexity of purity and inner product estimation. ArXiv e-prints, 2024.

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. doi: 10.1109/TIT.2017.2719044.

Masahito Hayashi. *Quantum Information Theory: Mathematical Foundation*. Cambridge University Press, 2016. doi: 10.1007/978-3-662-49725-8.

Masahito Hayashi. Measuring quantum relative entropy with finite-size effect. *Quantum*, 9:1725, 2025. doi: 10.22331/q-2025-05-05-1725.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

O. Hölder. Ueber einen mittelwerthabsatz. *Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen*, 1889:38–47, 1889. URL https://eudml.org/doc/180218.

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.

Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Maximum likelihood estimation of functionals of discrete distributions. *IEEE Transactions on Information Theory*, 63(10):6774–6798, 2017. doi: 10.1109/TIT.2017.2733537.

Sonika Johri, Damian S. Steiger, and Matthias Troyer. Entanglement spectroscopy on a quantum computer. *Physical Review B*, 96(19):195136, 2017. doi: 10.1103/PhysRevB.96.195136.

Yasuhito Kawano and Hiroshi Sekigawa. Quantum Fourier transform over symmetric groups — improved result. *Journal of Symbolic Computation*, 75:219–243, 2016. doi: 10.1016/j.jsc.2015.11.016.

Yuhan Liu and Jayadev Acharya. Quantum state testing with restricted measurements. ArXiv e-prints, 2024.

Yupan Liu and Qisheng Wang. On estimating the trace of quantum state powers. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 947–993, 2025. doi: 10.1137/1.9781611978322.28.

Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. In *Theory of Computing Library*, number 7 in Graduate Surveys, pages 1–81. University of Chicago, 2016. doi: 10.4086/toc.gs.2016.007.

Michael Nussbaum and Arleta Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *Annals of Statistics*, 37(2):1040–1057, 2009. doi: 10.1214/08-AOS593.

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

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

Ryan O'Donnell and John Wright. Quantum spectrum testing. *Communications in Mathematical Physics*, 387(1):1–75, 2021. doi: 10.1007/s00220-021-04180-1.

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.

Angelos Pelecanos, Xinyu Tan, Ewin Tang, and John Wright. Beating full state tomography for unentangled spectrum estimation. ArXiv e-prints, 2025.

Yihui Quek, Eneet Kaur, and Mark M. Wilde. Multivariate trace estimation in constant quantum depth. *Quantum*, 8:1220, 2024. doi: 10.22331/Q-2024-01-10-1220.

Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–562, 1961. URL https://static.renyi.hu/renyi_cikkek/1961_on_measures_of_entropy_and_information.pdf.

L. J. Roger. An extension of a certain theorem in inequalities. *Messenger of Mathematics*, 17(10):145–150, 1888. URL https://archive.org/details/messengermathem01unkngoog/page/n183/mode/1up?view=theater.

Myeongjin Shin, Junseo Lee, Seungwoo Lee, and Kabgyun Jeong. Resource-efficient algorithm for estimating the trace of quantum state powers. ArXiv e-prints, 2024.

Yiğit Subaşı, Lukasz Cincio, and Patrick J. Coles. Entanglement spectroscopy with a depth-two quantum circuit. *Journal of Physics A: Mathematical and Theoretical*, 52(4):044001, 2019. doi: 10.1088/1751-8121/aaf54d.

Constantino Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988. doi: 10.1007/BF01016429.

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 Annual ACM Symposium on Theory of Computing*, 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 Annual Symposium on Foundations of Computer Science*, pages 403–412, 2011b. doi: 10.1109/FOCS.2011.81.

Gregory Valiant and Paul Valiant. Estimating the unseen: improved estimators for entropy and other properties. *Journal of the ACM*, 64(6):37:1–37:41, 2017. doi: 10.1145/3125643.

S. J. van Enk and C. W. J. Beenakker. Measuring $\operatorname{Tr} \rho^n$ on single copies of ρ using random measurements. *Physical Review Letters*, 108(11):110503, 2012. doi: 10.1103/PhysRevLett.108.110503.

Qisheng Wang and Zhicheng Zhang. Quantum lower bounds by sample-to-query lifting. ArXiv e-prints, 2023.

Qisheng Wang and Zhicheng Zhang. Time-efficient quantum entropy estimator via samplizer. In *Proceedings of the 32nd Annual European Symposium on Algorithms*, pages 101:1–101:15, 2024a. doi: 10.4230/LIPIcs.ESA.2024.101.

Qisheng Wang and Zhicheng Zhang. Sample-optimal quantum estimators for pure-state trace distance and fidelity via samplizer. ArXiv e-prints, 2024b.

Qisheng Wang and Zhicheng Zhang. Fast quantum algorithms for trace distance estimation. *IEEE Transactions on Information Theory*, 70(4):2720–2733, 2024c. doi: 10.1109/TIT.2023.3321121.

Qisheng Wang, Ji Guan, Junyi Liu, Zhicheng Zhang, and Mingsheng Ying. New quantum algorithms for computing quantum entropies and distances. *IEEE Transactions on Information Theory*, 70(8):5653–5680, 2024a. doi: 10.1109/TIT.2024.3399014.

Xinzhao Wang, Shengyu Zhang, and Tongyang Li. A quantum algorithm framework for discrete probability distributions with applications to Rényi entropy estimation. *IEEE Transactions on Information Theory*, 70(5):3399–3426, 2024b. doi: 10.1109/TIT.2024.3382037.

John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139525343.

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.

Justin Yirka and Yiğit Subaşı. Qubit-efficient entanglement spectroscopy using qubit resets. *Quantum*, 5:535, 2021. doi: 10.22331/q-2021-09-02-535.

You Zhou and Zhenhuan Liu. A hybrid framework for estimating nonlinear functions of quantum states. *npj Quantum Information*, 10:62, 2024. doi: 10.1038/s41534-024-00846-5.

Appendix A. Proof of Lemma 13

To bound the success probability, we need Hoeffding's inequality.

Theorem 21 (Hoeffding's inequality, (Hoeffding, 1963, Theorem 2)) *Let $X_1, X_2, \ldots, X_n$ be independent and identical random variables with $X_j \in [0, 1]$ for all $1 \leq j \leq n$. Then,*

$$\Pr\left[\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]\right| \leq t\right] \geq 1 - 2\exp\bigl(-2nt^2\bigr).$$

Proof [Proof of Lemma 13] We present a formal description of our approach in Algorithm 1. In the following proof, all expectations are computed over $\lambda \sim \mathrm{SW}^n(\alpha)$. Let $\underline{\lambda}_j = \lambda_j/n$. By Lemma 11, we have

$$\mathbb{E}\bigl[(\underline{\lambda}_j - \alpha_j)^2\bigr] \leq \frac{c}{n},$$

for some constant $c > 0$. Therefore,

$$\begin{aligned} \Pr\left[|\underline{\lambda}_j - \alpha_j| \geq 2\sqrt{\frac{c}{n}}\right] \cdot 4 \cdot \frac{c}{n} &\leq \mathbb{E}\bigl[(\underline{\lambda}_j - \alpha_j)^2\bigr] \qquad (17) \\ &\leq \frac{c}{n}, \end{aligned}$$

where Equation (17) is by Markov's inequality that $\Pr[|X| \geq a] \cdot a^k \leq \mathbb{E}[|X|^k]$ for any random variable X, integer $k \geq 1$, and $a > 0$. This implies

$$\Pr\left[|\underline{\lambda}_j - \alpha_j| \geq 2\sqrt{\frac{c}{n}}\right] \leq \frac{1}{4}.$$

Let $k \geq 1$ be an odd integer. Now we draw k independent samples $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(k)}$ from $\mathrm{SW}^n(\alpha)$, and for each $j \in [d]$ let

$$\hat{\alpha}_j = \mathrm{median}\Bigl\{\underline{\lambda}_j^{(1)}, \underline{\lambda}_j^{(2)}, \ldots, \underline{\lambda}_j^{(k)}\Bigr\}.$$

Let $X_j^{(l)} \in \{0, 1\}$ be a random variable such that $X_j^{(l)} = 1$ if $|\underline{\lambda}_j^{(l)} - \alpha_j| \geq 2\sqrt{c/n}$ and 0 otherwise. By Hoeffding's inequality (Theorem 21) with $t = 1/12$, we have

$$\Pr\left[\left|\frac{1}{k}\sum_{l=1}^{k} X_j^{(l)} - \mathbb{E}\Bigl[X_j^{(1)}\Bigr]\right| \leq \frac{1}{12}\right] \geq 1 - 2\exp\left(-\frac{k}{72}\right).$$

Note that $\mathbb{E}[X_j^{(1)}] \leq 1/4$, then

$$\Pr\left[\frac{1}{k}\sum_{l=1}^{k} X_j^{(l)} \leq \frac{1}{3}\right] \geq 1 - 2\exp\left(-\frac{k}{72}\right),$$

which means that $\hat{\alpha}_j$, the median of $\underline{\lambda}_j^{(1)}, \underline{\lambda}_j^{(2)}, \ldots, \underline{\lambda}_j^{(k)}$, satisfies $|\hat{\alpha}_j - \alpha_j| \leq 2\sqrt{c/n}$ with probability

$$\Pr\left[|\hat{\alpha}_j - \alpha_j| \leq 2\sqrt{\frac{c}{n}}\right] \geq 1 - 2\exp\left(-\frac{k}{72}\right).$$

By taking $n = \lceil 4c/\varepsilon^2 \rceil$ and $k = \lceil 72 \ln(2/\delta) \rceil$, we have

$$\Pr[|\hat{\alpha}_j - \alpha_j| \leq \varepsilon] \geq 1 - \delta,$$

which uses $nk = O(\log(1/\delta)/\varepsilon^2)$ samples of ρ. ■

Appendix B. Proof of Lemma 19

Proof Note that if $x_i \geq x_j$ and $0 \leq \Delta \leq x_j$, then it is easy to verify that

$$x_i^q + x_j^q \leq (x_i + \Delta)^q + (x_j - \Delta)^q.$$

For any sequence $x_m \geq x_{m+1} \cdots \geq x_N \geq 0$, we define a new sequence by the following process:

1. Find the smallest index j such that $x_j < x_m$, and then find the largest index k such that $k > j$ and $x_k > 0$. If there are no such j, k, then do nothing.
2. Upon the success of finding j, k, we define the new sequence by $x_i' = x_i$ for all $i \neq j, k$ and

   $$x_j' = x_j + \Delta, \quad x_k' = x_k - \Delta,$$

   where $\Delta = \min\{x_m - x_j, x_k\}$.

It is obvious that

$$x_{m+1}^q + \cdots + x_N^q \leq (x_{m+1}')^q + \cdots + (x_N')^q.$$

Starting from a sequence $x_m \geq x_{m+2} \cdots \geq x_N$, we define $A = \sum_{i=m+1}^N x_i$. Then, we iteratively apply this process and finally get a sequence like

$$x_m, \underbrace{x_m, x_m, \ldots, x_m}_{l}, y,$$

where $l = \lfloor A/x_m \rfloor$ and $y = A - l \cdot x_m$. Therefore

$$\begin{aligned}
\sum_{i=m+1}^N x_i^q &\leq l \cdot x_m^q + y^q \\
&= x_{m+1}^q \left(l + \left(\frac{A}{x_m} - l \right)^q \right) \\
&\leq x_m^q \cdot \frac{A}{x_m} && (18) \\
&\leq x_m^{q-1}, && (19)
\end{aligned}$$

where Equation (18) is because $A/x_m - l < 1$. Then, by noting that $x_m \leq 1/m$, we have

$$(19) \leq \frac{1}{m^{q-1}}.$$

■

Appendix C. Simple lower bounds by quantum state discrimination

Theorem 22 *For any constant $q > 1$, any quantum estimator to additive error ε for $\mathrm{tr}(\rho^q)$ requires sample complexity $\Omega(1/\varepsilon^2)$.*

Proof Consider the problem of distinguishing the two quantum states $\rho_\pm$, where $\rho_\pm = \left(\frac{2}{3} \pm \varepsilon\right)|0\rangle\langle 0| + \left(\frac{1}{3} \mp \varepsilon\right)|1\rangle\langle 1|$. Then,

$$\mathrm{tr}(\rho_\pm^q) = \left(\frac{2}{3} \pm \varepsilon\right)^q + \left(\frac{1}{3} \mp \varepsilon\right)^q.$$

$$\mathrm{tr}(\rho_+^q) - \mathrm{tr}(\rho_-^q) = \left(\left(\frac{2}{3} + \varepsilon\right)^q - \left(\frac{2}{3} - \varepsilon\right)^q\right) + \left(\left(\frac{1}{3} - \varepsilon\right)^q - \left(\frac{1}{3} + \varepsilon\right)^q\right).$$

By the direct calculation that

$$\lim_{\varepsilon \to 0} \frac{\mathrm{tr}(\rho_+^q) - \mathrm{tr}(\rho_-^q)}{\varepsilon} = 2q\left(\left(\frac{2}{3}\right)^{q-1} - \left(\frac{1}{3}\right)^{q-1}\right) = \Theta(1),$$

we conclude that $\mathrm{tr}(\rho_+^q) - \mathrm{tr}(\rho_-^q) = \Theta(\varepsilon)$. Therefore, any quantum estimator for $\mathrm{tr}(\rho^q)$ to additive error $\Theta(\varepsilon)$ can be used to distinguish ρ_+ and ρ_-. On the other hand, if the quantum estimator for $\mathrm{tr}(\rho^q)$ to additive error ε has sample complexity S, then $S \geq \Omega(1/\gamma)$. A direct calculation shows that the infidelity

$$\gamma = 1 - \mathrm{F}(\rho_+, \rho_-) = 1 - \left(\sqrt{\frac{4}{9} - \varepsilon^2} + \sqrt{\frac{1}{9} - \varepsilon^2}\right) = \Theta(\varepsilon^2).$$

By Fact 4, we have $S = \Omega(1/\varepsilon^2)$. ∎