Trace Estimation of Quantum State Powers: Sample Complexity and Computational Hardness

Kean Chen * Yupan Liu † Qisheng Wang ‡

Abstract

As often emerges in various basic quantum properties such as Rényi and Tsallis entropies, 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, even for (possibly) non-integer $q > 1$, $\mathrm{tr}(\rho^q)$ can be estimated to within additive error ε using a dimension-independent (and also rank-independent) sample complexity of $\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$,[1] together with a lower bound of $\Omega(1/\varepsilon)$. In addition, combining this result with subsequent work of Liu (STACS 2026) shows that the corresponding promise problem is BQP-complete. In this paper, we significantly improve and extend the sample complexity bounds for this problem. Furthermore, we show that for $0 < q < 1$, the problem does not admit an efficient estimator unless $\mathsf{BQP} = \mathsf{NIQSZK}$, which is considered highly unlikely. In particular, we have the following results.

- For $q > 2$, we settle the sample complexity with matching upper and lower bounds $\widetilde{\Theta}(1/\varepsilon^2)$.
- For $1 < q < 2$, we obtain an upper bound of $\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$, with a lower bound of $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1},2\}})$ for dimension-independent (in fact, rank-independent) estimators.
- For $0 < q < 1$, we obtain an upper bound of $O((d/\varepsilon)^{\frac{2}{q}})$, with a lower bound of $\Omega((d/\varepsilon)^{\frac{1}{q}})$ for d-dimensional states (in fact, both bounds can be naturally refined to depend on the rank rather than the dimension). Accordingly, the corresponding promise problem is NIQSZK-hard, which is in sharp contrast to the case of $q > 1$.

Technically, 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.

A preliminary version of this paper [CW25] was presented at the 38th Conference on Learning Theory (COLT 2025).

*keanchen.gan@gmail.com
†yupan.liu@epfl.ch
‡QishengWang1994@gmail.com

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

Contents

1 Introduction

Testing the properties of quantum states is a fundamental problem in the field of quantum property testing [MdW16], 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 [OW21], testing the spectrum was extensively studied, with several significant applications such as mixedness testing and rank testing. In [OW17], 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 [AISW20, BMW16, WZ25b].

The traces of quantum state powers, $\operatorname{tr}(\rho^q)$, of a quantum state ρ are one of the simplest functionals of quantum states. The quantity $\operatorname{tr}(\rho^q)$ has connections to the Rényi entropy $\mathrm{S}_q^{\mathrm{R}}(\rho) = \frac{1}{1-q}\ln(\operatorname{tr}(\rho^q))$ [Rén61] and the Tsallis entropy $\mathrm{S}_q^{\mathrm{T}}(\rho) = \frac{1}{1-q}(\operatorname{tr}(\rho^q) - 1)$ [Tsa88]. The estimation of $\operatorname{tr}(\rho^q)$ 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 systems [Bec01, Bec02]. 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 [LW25]. In particular, $\operatorname{tr}(\rho^2)$ refers to the purity of ρ, and it is well-known that the purity $\operatorname{tr}(\rho^2)$ can be estimated to within additive error ε using $O(1/\varepsilon^2)$ samples of ρ via the SWAP test [BCWdW01]. For the case of constant integer $q \geq 2$, $\operatorname{tr}(\rho^q)$ can be estimated using $O(1/\varepsilon^2)$ samples of ρ via the Shift test proposed in [EAO+02], generalizing the SWAP test. For non-integer $q > 0$ and $q \neq 1$, the estimation of $\operatorname{tr}(\rho^q)$ was considered in [WGL+24] with the corresponding quantum algorithms presented with time complexity $\operatorname{poly}(r, 1/\varepsilon)$,[2] where r is the rank of ρ. Recently in [LW25], it was discovered that for every non-integer $q > 1$, $\operatorname{tr}(\rho^q)$ 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 improves the results in [WGL+24] and the results implied by other works [AISW20, WZL24, WZ25b] on Rényi entropy estimation. However, the sample complexity in [LW25] is far from being optimal, as only a lower bound of $\Omega(1/\varepsilon)$ on the sample complexity of estimating $\operatorname{tr}(\rho^q)$ for non-integer $q > 1$ was known in [LW25, 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 $\operatorname{tr}(\rho^2)$ (see [CWLY23, Theorem 5] and [GHYZ24, Lemma 3]).

Beyond prior work on the query and sample complexities of estimating the von Neumann entropy and the Tsallis entropy $\mathrm{S}_q^{\mathrm{T}}(\rho)$, it is natural to consider the corresponding promise problem that distinguishes whether $\mathrm{S}_q^{\mathrm{T}}(\rho) \geq \tau_0$ or $\mathrm{S}_q^{\mathrm{T}}(\rho) \leq \tau_1$ for constants $\tau_0 > \tau_1$. In the von Neumann entropy case ($q = 1$), this problem is NIQSZK-complete [Kob03, BST10, CCKV08], a class that originally arose in the study of (non-interactive) quantum statistical zero-knowledge [Wat02, Kob03]. Regardless of its origin, this result implies that estimating the von Neumann entropy is computationally equivalent to distinguishing any efficiently preparable quantum state from the maximally mixed state with respect to the trace distance.[3] In contrast, for constant $q > 1$, the complexity landscape changes in a manner consistent with the known efficient estimators: it was shown in [LW25] that the regime $q \in (1, 2]$ is BQP-complete, while the regime $q > 2$ lies in BQP. An open problem posed in the conference version of this work [CW25], asking whether constant $q > 2$ is also BQP-hard,

[2] In [WGL+24], their main results only consider the quantum query complexity, as they assume access to the state-preparation circuit of ρ. Nevertheless, their results also imply a sample complexity of $\operatorname{poly}(r, 1/\varepsilon)$ (with a polynomial overhead compared to the corresponding query complexity) using the techniques in [GP22], as noted in [WGL+24, Footnote 2].

[3] While our work focuses on the case where q is a real number, it is noteworthy that the NIQSZK-hardness result extends to the regime where q is slightly above 1, specifically $1 \leq q \leq 1 + \frac{1}{n}$, as proven in [LW25].

was resolved affirmatively in [Liu26]. Together with [LW25], this establishes that the associated problem is BQP-complete for all constant $q > 1$.

In this paper, we further investigate the sample complexity and the computational hardness of estimating $\mathrm{tr}(\rho^q)$. For $q > 1$, we achieve significant improvements over the prior results [LW25] 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. For $0 < q < 1$, we complement the literature by providing estimators and showing that the corresponding promise problem is NIQSZK-hard, which is in sharp contrast to the BQP-completeness for $q > 1$ due to [LW25, Liu26]. Our results are collected in Section 1.1. In addition, it is noteworthy that our techniques are conceptually and technically different from those in [LW25]. In comparison, our estimator is based on weak Schur sampling [CHW07] while the estimator in [LW25] is based on quantum singular value transformation [GSLW19] and samplizer [WZ25a, WZ25b]. Concerning the computational hardness, while our result follows the framework of [CCKV08, KLN19] at a conceptual level, our main technical contribution is a new inequality bounding the q-Tsallis entropy for $0 < q < 1$ in terms of its closeness to the uniform distribution, which is of independent interest. For more details, see Section 1.2.

1.1 Main Results

To illustrate our results, we present them in three parts separately: $q > 2$, $1 < q < 2$, and $0 < q < 1$.

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.

Theorem 1.1 (Optimal estimator for $q > 2$, informal version of Theorems 3.1 and 3.2)**.** *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 a lower bound of $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1},2\}})$.

Theorem 1.2 (Improved estimator for $1 < q < 2$, informal version of Theorems 3.2, 4.1 and 4.3)**.** *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 $\mathrm{tr}(\rho^q)$ 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.*

Both Theorems 1.1 and 1.2 improve the prior best upper bound $\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$ and lower bound $\Omega(1/\varepsilon)$ in [LW25]. It is also noted that Theorem 1.1 gives a matching lower bound of $\Omega(1/\varepsilon^2)$ on the sample complexity of estimating $\mathrm{tr}(\rho^q)$ for every integer $q \geq 3$, implying that the Shift test in [EAO+02] is sample-optimal to estimate $\mathrm{tr}(\rho^q)$ to within an additive error, generalizing the lower bounds in [CWLY23, GHYZ24] for the optimality of the SWAP test [BCWdW01] to estimate $\mathrm{tr}(\rho^2)$.

The case of $0 < q < 1$. For $0 < q < 1$, we provide a quantum estimator with sample complexity $O(d^{2/q}/\varepsilon^{2/q})$, only with room for quadratic improvements due to a lower bound of $\Omega(d^{1/q}/\varepsilon^{1/q})$.

Theorem 1.3 (Estimator for $0 < q < 1$, informal version of Theorems 5.1 and 5.2)**.** *For every $0 < q < 1$, it is sufficient to use $O(d^{2/q}/\varepsilon^{2/q})$ samples of the d-dimensional quantum state ρ to estimate $\mathrm{tr}(\rho^q)$ to within additive error ε. On the other hand, $\Omega(d^{1/q}/\varepsilon^{1/q})$ samples of ρ are necessary.*

In Theorem 1.3, if the rank of ρ is known to be at most r in advance, then we can replace d in our upper and lower bounds with r. It is noteworthy that Theorem 1.3 also implies an estimator for the Rényi entropy $\mathrm{S}_q^{\mathrm{R}}(\rho)$ for $0 < q < 1$ with sample complexity $O(d^{2/q}/\varepsilon^{2/q})$,[4] thus giving an alternative proof of [AISW20, Theorem 4].

Remark 1.1 (Time efficiency of the estimators in Theorems 1.1 to 1.3)**.** *The estimators in Theorems 1.1 and 1.2 can actually be implemented with quantum time complexity* $\mathrm{poly}(\log(d), 1/\varepsilon)$ *for any constant* $q > 1$. *This is because, in their implementations (Algorithms 2 and 3), we only need the first* m *entries of the output of the quantum spectrum estimation (Algorithm 1) with* $m \leq O(1/\varepsilon^{\max\{1, \frac{1}{q-1}\}})$. *On the other hand, Algorithm 1 uses* $\mathsf{n} = \widetilde{O}(1/\varepsilon^{\max\{\frac{2}{q-1}, 2\}})$ *samples of* ρ *and can be implemented with quantum time complexity* $O(\mathsf{n}^3 \operatorname{polylog}(\mathsf{n}, d)) = \widetilde{O}(1/\varepsilon^{\max\{\frac{6}{q-1}, 6\}}) \cdot \operatorname{polylog}(d)$ *by weak Schur sampling [CHW07].*[5] *Similarly, the estimator in Theorem 1.3 can be implemented in Algorithm 4, using* $\mathsf{n} = O((d/\varepsilon)^{2/q})$ *samples of* ρ *and thus with quantum time complexity* $O(\mathsf{n}^3 \operatorname{polylog}(\mathsf{n}, d)) = \widetilde{O}((d/\varepsilon)^{6/q}) \cdot \operatorname{polylog}(d)$.

We summarize the developments for the sample complexity of estimating $\operatorname{tr}(\rho^q)$ in Table 1.

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

$q \geq 2$	$1 < q < 2$	$0 < q < 1$	References
$O(1/\varepsilon^2)$, $q \in \mathbb{N}$	/	/	[BCWdW01, EAO+02]
$\Omega(1/\varepsilon^2)$, $q = 2$	/	/	[CWLY23, GHYZ24]
$\mathrm{poly}(d, 1/\varepsilon)$			[AISW20, WGL+24, WZL24, WZ25b]
$\widetilde{O}(1/\varepsilon^{3+\frac{2}{q-1}})$, $\Omega(1/\varepsilon)$		/	[LW25]
$\widetilde{\Theta}(1/\varepsilon^2)$	$\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$ $\Omega(1/\varepsilon^{\max\{\frac{1}{q-1}, 2\}})$	$\widetilde{O}(d^{2/q}/\varepsilon^{2/q})$ $\Omega(d^{1/q}/\varepsilon^{1/q})$	This Work

Moreover, from a complexity-theoretic perspective, we establish the computational hardness of estimating $\operatorname{tr}(\rho^q)$ for $0 < q < 1$ as follows.

Theorem 1.4 (Informal version of Theorem 5.5)**.** *For every* $q \in (0, 1)$, *the promise problem of estimating* $\operatorname{tr}(\rho^q)$, $\textsc{TsallisQEA}_q$, *is* NIQSZK*-hard.*

It follows that an efficient estimator cannot exist unless NIQSZK collapses to BQP, which is highly unlikely and thus consistent with the qualitative change in the sample complexity bounds across different regimes of q in Table 1; in particular, an efficient estimator exists only in those regimes of q for which the corresponding promise problem is BQP-complete. In summary, we collect the computational hardness results for $\textsc{TsallisQEA}_q$ in Table 2.

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 [CHW07,

[4] Any estimator for the Tsallis entropy $\mathrm{S}_q^{\mathrm{T}}(\rho)$ is an estimator for the Rényi entropy $\mathrm{S}_q^{\mathrm{R}}(\rho)$ to the same additive error for $0 < q < 1$.

[5] This quantum time complexity was noted in [WZ24a, WZ25b, Hay25]. This is achieved by using the implementation of weak Schur sampling introduced in [MdW16, Section 4.2.2], equipped with the quantum Fourier transform over symmetric groups in [KS16].

Table 2: Computational hardness of $\textsc{TsallisQEA}_q$.

$q > 2$	$1 < q \leq 2$	$q = 1$ (von Neumann)	$0 < q < 1$
BQP-complete [LW25, Liu26]	BQP-complete [LW25]	NIQSZK-complete [BST10, CCKV08]	NIQSZK-hard This Work

MdW16, OW21] 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 [AISW20, BMW16].

However, directly using the plug-in estimator with current error bounds for weak Schur sampling in [OW17] 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 [JVHW15, JVHW17] 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 $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d)$ of ρ to certain precision with $\hat{\alpha}_1 \geq \hat{\alpha}_2 \geq \cdots \geq \hat{\alpha}_d$ (with $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ 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 [AISW20, BMW16], our non-plug-in construction can be shown to achieve optimal sample complexity only up to a logarithmic factor (see Sections 3.1 and 4.1 for more details). As a result, we obtain sample upper bounds $\widetilde{O}(1/\varepsilon^2)$ for $q > 2$, $\widetilde{O}(1/\varepsilon^{\frac{2}{q-1}})$ for $1 < q < 2$, and $\widetilde{O}(d^{2/q}/\varepsilon^{2/q})$ for $0 < q < 1$. 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 (5)), causing q to disappear from the exponent.

Lower bounds. The 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 [CWLY23, GHYZ24].

The lower bound $\Omega(1/\varepsilon^{\frac{1}{q-1}})$ for $1 < q < 2$ is obtained by reducing a state 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 [CHW07], which demands a linear scaling with the dimensions. With carefully chosen dimension parameters, we can obtain our lower bound.

The lower bound $\Omega(d^{1/q}/\varepsilon^{1/q})$ for $0 < q < 1$ can also be established via a reduction from a state discrimination task over ensembles of quantum states. Compared to the previous construction, we introduce an additional parameter Δ that controls the frequency of "valid" samples: among n samples, only $O(\Delta n)$ of them (with high probability) contribute to the discrimination task. These "valid" samples exhibit unitary invariance and induce specific Schur–Weyl distributions. We can then upper bound the relevant trace distance by using the Plancherel distribution as an intermediate tool [CHW07].

Computational hardness for $0 < q < 1$. Our NIQSZK-hardness results conceptually follow the framework of [CCKV08, KLN19], which treats the case $q = 1$ corresponding to the von Neumann entropy. At a high level, this framework reduces QSCMM, the problem of distinguishing an arbitrary efficiently preparable n-qubit state ρ from the maximally mixed state $(I/2)^{\otimes n}$ with respect to the trace distance, to the promise problem $\textsc{TsallisQEA}_q$ of estimating the q-Tsallis entropy of ρ. Because ρ and $(I/2)^{\otimes n}$ are simultaneously diagonalizable, the reduction relies on an inequality relating the q-Tsallis entropy of the eigenvalue distribution of ρ to its total variation distance from the uniform distribution.

Proving such an inequality leads to an optimization problem over the eigenvalue distribution. For the regime $1 \leq q \leq 1 + \frac{1}{n}$, the (possibly non-convex) optimization problems arising in the proofs of [KLN19, LW25] admit explicit solutions.[6] In fact, a direct analog of this approach is available only for $q = 1/2$. In contrast, for the general regime $0 < q < 1$, obtaining an explicit solution is often challenging due to the non-convex nature of the optimization problem. Instead, the q-Tsallis entropy is upper bounded by analyzing a truncated Taylor expansion of the objective function together with a Lagrange remainder term, yielding a weaker bound (see Theorem 5.6) that nevertheless suffices to establish the reduction $\text{QSCMM} \leq \textsc{TsallisQEA}_q$ and hence the NIQSZK-hardness.

1.3 Related Work

After the work of [BCWdW01, EAO+02], there have been a series of subsequent work focusing on the estimation of $\operatorname{tr}(\rho^q)$ for integer $q \geq 2$ [Bru04, vEB12, JST17, SCC19, YS21, QKW24, ZL24, SLLJ25, YLLW24, CWYZ25, Wan25, SJW+25]. The quantum query complexity of entropy estimation has also been extensively studied in the literature, including the von Neumann entropy [GL20, GHS21, WGL+24] and Rényi entropy [SH21, WGL+24, WZL24]. 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 [AK01] for integer $q \geq 2$, and later in [JVHW15, JVHW17] for non-integer q; its estimation to a multiplicative error was studied in [AOST17] for Rényi entropy estimation. In addition, Shannon entropy estimation was studied in [Pan03, Pan04, VV11a, VV11b, VV17, WY16].

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 [CHW07], 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 [OW21] and to quantum state certification in [BOW19]. In addition, optimal estimators are known for Rényi entropy of integer order [AISW20], and the closeness (trace distance and fidelity) between pure quantum states [WZ24b]. A distributed optimal estimator was known for the inner product of quantum states [ALL22]. Estimators and testers with incoherent measurements are also known

[6] As observed in [KLN19], when $q = 1$, the upper bound on the Shannon entropy in this setting follows from Vajda's inequality [Vaj70].

for purity [CCHL21, GHYZ24], unitarity [CCHL21, CWLY23], certification [CHLL22, LA24], and $\mathrm{tr}(\rho^q)$ for integer q (further used for spectrum estimation) [PTTW26]. In addition to those that were known to be optimal, there are also estimators for entropy [AISW20, BMW16, WZ25b, LW25], relative entropy [Hay25], fidelity [GP22], and trace distance [WZ24a].

1.4 Discussion

In this paper, we presented quantum estimators for estimating $\mathrm{tr}(\rho^q)$ 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 [MdW16, Lemma 20]), in addition to the (plug-in) optimal estimator for Rényi entropy of integer order [AISW20], the optimal testers for mixedness [OW21] and quantum state certification [BOW19], and the optimal learners for full tomography [HHJ+17, OW16]. 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 $\mathrm{tr}(\rho^q)$ with restricted measurements?

2 Preliminaries

2.1 Notations

In our quantum algorithms for the cases of $q > 2$ and $1 < q < 2$, 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 2.5. 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. For the case of $0 < q < 1$, we use the parameter η, which specifies the precision for estimating the spectrum of ρ in the total variation distance. Throughout this paper, we assume that ρ has the spectrum decomposition:

$$\rho = \sum_{j=1}^{d} \alpha_j |\psi_j\rangle\langle\psi_j|,$$

where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_d \geq 0$ with $\sum_{j=1}^{d} \alpha_j = 1$ and $\{|\psi_j\rangle\}$ is an orthonormal basis.

In addition, we adopt the notion of the *q-logarithm* $\ln_q(x) \coloneqq \frac{x^{1-q}-1}{1-q}$, which converges to the natural logarithm $\ln(x)$ as $q \to 1$. This notion arises naturally in the definition of the quantum q-Tsallis entropy, as reflected by the alternative expression $\mathrm{S}_q^{\mathrm{T}}(\rho) = -\mathrm{tr}(\rho^q \ln_q(\rho))$.

For our purpose, we also need the following inequalities.

Fact 2.1. *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 $f(x) = x^\alpha$. □

Fact 2.2. *For $0 \leq x \leq y \leq 1$ and $0 < s < 1$ we have $y^s - x^s \leq (y-x)^s$.*

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

Fact 2.3. *For $0 < s < 1$ and $x_i \geq 0$ for all $1 \leq i \leq k$, we have*

$$\sum_{i=1}^{k} x_i^s \leq k^{1-s} \cdot \left(\sum_{i=1}^{k} x_i\right)^s.$$

Proof. Let $y_i = x_i^s$. By Roger–Hölder's inequality [Rog88, Höl89],

$$\sum_{i=1}^{k} x_i^s = \sum_{i=1}^{k} y_i \leq \left(\sum_{i=1}^{k} 1^{\frac{1}{1-s}}\right)^{1-s} \left(\sum_{i=1}^{k} y_i^{\frac{1}{s}}\right)^s = k^{1-s} \cdot \left(\sum_{i=1}^{k} x_i\right)^s.$$

□

2.2 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

$$\mathrm{T}(\rho_0, \rho_1) \coloneqq \frac{1}{2}\|\rho_0 - \rho_1\|_1 = \frac{1}{2}\operatorname{tr}(|\rho_0 - \rho_1|).$$

The fidelity between two quantum states ρ_0 and ρ_1 is defined by

$$\mathrm{F}(\rho_0, \rho_1) \coloneqq \operatorname{tr}\left(\sqrt{\sqrt{\rho_1}\rho_0\sqrt{\rho_1}}\right).$$

To discriminate two quantum states, we include the following well-known results. The following theorem can be found in [Wil13, Section 9.1.4], [Hay16, Lemma 3.2], and [Wat18, Theorem 3.4].

Theorem 2.4 (Quantum state discrimination, cf. [Wil13, Section 9.1.4], [Hay16, Lemma 3.2], and [Wat18, Theorem 3.4])**.** *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*

$$\frac{1}{2}\operatorname{tr}(\Lambda_0\rho_0) + \frac{1}{2}\operatorname{tr}(\Lambda_1\rho_1) \leq \frac{1}{2}\left(1 + \frac{1}{2}\|\rho_0 - \rho_1\|_1\right).$$

The following fact was noted in [HHJ+17, Section 1], which is closely related to the quantum Chernoff bound [NS09, ACMT+07].

Fact 2.5. *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.*

2.3 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 $\mathcal{H}$.[7] We also call $\mu(g)$ the

[7]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.

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 a 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.

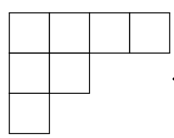
Fact 2.6 (Schur's Lemma, see, e.g. [EGH+11, Proposition 2.3.9])**.** *Let $\mathcal{H}_1, \mathcal{H}_2$ be irreducible representations of a group G. If $F : \mathcal{H}_1 \to \mathcal{H}_2$ is a non-zero homomorphism of representations, then F is an isomorphism.*

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

Corollary 2.7. *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.*

2.3.1 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 $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$. For example, the Young diagram with 8 boxes and 3 rows, identified by the partition $(4, 2, 1)$ is:



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 commute with each other, 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.

Fact 2.8 (Schur–Weyl duality [FH13, EGH+11])**.**

$$(\mathbb{C}^d)^{\otimes n} \overset{\mathfrak{S}_n \times \mathbb{U}_d}{\cong} \bigoplus_{\lambda \vdash n} \mathcal{P}_\lambda \otimes \mathcal{Q}_\lambda^d,$$

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 $\lambda \vdash n$.[8]

[8]Note that if the Young diagram λ has more than d rows, then $\mathcal{Q}_\lambda^d = 0$.

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 2.1. *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 2.9. *$X^{\otimes n}$ 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 2.1).

Furthermore, it is known that $\mathrm{tr}(\mathsf{q}_\lambda(X)) = s_\lambda(\alpha)$, where s_λ is the *Schur polynomial* [FH13] indexed by λ and $\alpha = (\alpha_1, \ldots, \alpha_d)$ are the eigenvalues of X.

2.4 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 $\rho^{\otimes n}$,[9] 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:

$$\overline{M}_i = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \pi \mathop{\mathbb{E}}_{U \in \mathbb{U}_d} \left[U^{\otimes n} M_i U^{\dagger \otimes n} \right] \pi^\dagger.$$

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., [MdW16, Hay25]).

Fact 2.10. *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 2.8) and Corollary 2.7, 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 2.9, we can see that the estimator $\{\overline{M}_i\}$ applied on $\rho^{\otimes n}$ is equivalent to

[9] Here, we assume the POVM is discrete, the continuous case can be treated similarly.

1. sample a $\lambda \vdash n$ from the distribution $\{\mathrm{tr}(I_{\mathcal{P}_\lambda} \otimes \mathfrak{q}_\lambda(\rho))\}_\lambda = \{\dim(\mathcal{P}_\lambda) \cdot s_\lambda(\alpha)\}_\lambda$, where s_λ is the Schur polynomial and $\alpha = (\alpha_1, \ldots, \alpha_d)$ are the eigenvalues of ρ such that $\alpha_1 \geq \cdots \geq \alpha_d$.
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* [CHW07]. In step 1, the distribution $\{\dim(\mathcal{P}_\lambda) \cdot s_\lambda(\alpha)\}_\lambda$ is referred to as the *Schur–Weyl distribution* [OW17] 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 results.

Lemma 2.11 ([OW16, Theorem 1.7])**.** *It holds that*

$$\mathop{\mathbb{E}}_{\lambda \sim \mathrm{SW}^n(\alpha)}\left[\frac{1}{2}\sum_{j \in [d]}\left|\frac{\lambda_j}{n} - \alpha_j\right|\right] \leq \frac{1.92d + 0.5}{\sqrt{n}}.$$

Lemma 2.12 (Adapted from [OW17, Theorem 1.5])**.** *For $j \in [d]$,*

$$\mathop{\mathbb{E}}_{\lambda \sim \mathrm{SW}^n(\alpha)}\left[(\lambda_j - \alpha_j n)^2\right] \leq O(n).$$

We use SW^n_d to denote $\mathrm{SW}^n(\alpha)$ when $\alpha = (1/d, \ldots, 1/d)$,[10] 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 2.13 ([CHW07, Lemma 6])**.** *If $n \leq d$, then*

$$\|\mathrm{SW}^n_d - \mathrm{Planch}(n)\|_1 \leq \sqrt{2}\frac{n}{d}.$$

2.5 Quantum spectrum estimation with entry-wise bounds

Efficient approaches to quantum spectrum estimation were given in [OW16] in the ℓ_1 and ℓ_2 distances and in [OW17] in the Hellinger-squared distance, chi-squared divergence, and Kullback–Leibler (KL) divergence. In this section, we provide an efficient approach to quantum spectrum estimation with entry-wise bounds in Algorithm 1 based on the results of [OW17], which will be used as a subroutine in our estimators for $\mathrm{tr}(\rho^q)$.

For our purposes, we consider the case where $k > 1$, which is useful in our estimation algorithms for the case of $q > 1$ so that we can remove the dependence on d in the complexity.

Lemma 2.14 (Quantum spectrum estimation with entry-wise bounds)**.** *For every $\varepsilon, \delta \in (0, 1)$, the process* `SpectrumEstimation`$(\rho, n, k)$ *with $n = \Theta(1/\varepsilon^2)$ and $k = \Theta(\log(1/\delta))$ uses $nk = O(\log(1/\delta)/\varepsilon^2)$ samples of ρ and returns 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 $|\hat{\alpha}_j - \alpha_j| \leq \varepsilon$.*

[10] In some papers, SW^n_d is also called the Schur–Weyl distribution [OW21] or simply the Schur distribution [CHW07].

Algorithm 1 `SpectrumEstimation`(ρ, n, k)

Input: Sample access to a d-dimensional mixed quantum state ρ; integers $n, k \geq 1$.
Output: A d-dimensional vector $\hat{\alpha} \in \mathbb{R}^d$.

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

We also use the special case of Algorithm 1 with $n = \Theta(d/\varepsilon^2)$ and $k = 1$, which is actually the standard result on quantum state tomography in [OW16].

Lemma 2.15 (Quantum spectrum estimation in total variation distance, adapted from [OW16, Theorem 1.7])**.** *For every $\varepsilon \in (0, 1)$, the process* `SpectrumEstimation`(ρ, n, k) *with $n = \Theta(d/\varepsilon^2)$ and $k = 1$ uses $nk = O(d/\varepsilon^2)$ samples of ρ and returns a sequence of random variables $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d) \in \mathbb{R}^d$ such that it holds with probability at least $2/3$ that*

$$\sum_{j=1}^{d} |\hat{\alpha}_j - \alpha_j| \leq \varepsilon.$$

We also need Hoeffding's inequality.

Theorem 2.16 (Hoeffding's inequality, [Hoe63, 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,*

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

Proof of Lemma 2.14. In this proof, all expectations are computed over $\lambda \sim \mathrm{SW}^n(\alpha)$. Let $\underline{\lambda}_j = \lambda_j/n$. By Lemma 2.12, we have

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

for some constant $c > 0$. Therefore,

$$\begin{aligned} \Pr\left[|\underline{\lambda}_j - \alpha_j| \geq 2\sqrt{\frac{c}{n}}\right] \cdot 4 \cdot \frac{c}{n} &\leq \mathbb{E}\left[(\underline{\lambda}_j - \alpha_j)^2\right] \qquad (1) \\ &\leq \frac{c}{n}, \end{aligned}$$

where Equation (1) 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

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

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

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

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 2.16) with $t = 1/12$, we have

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

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 that for any $j \in [d]$,

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

The whole process uses $nk = O(\log(1/\delta)/\varepsilon^2)$ samples of ρ. □

2.6 Computational hardness of QSCMM

We begin by defining the promise problem QSCMM:

Definition 2.17 (Quantum State Closeness to Maximally Mixed State, QSCMM$[\beta, \alpha]$, adapted from [Kob03, Section 3]). *Let Q be a quantum circuit acting on m qubits with n specified output qubits, satisfying $m \geq 2n$.*[11] *Let ρ denote the quantum state obtained by applying Q to the initial state $|0\rangle^{\otimes m}$ and tracing out the non-output qubits. Let $\alpha(n)$ and $\beta(n)$ be efficiently computable functions. The task is to decide whether the following holds:*

- **Yes:** $\mathrm{T}(\rho, (I/2)^{\otimes n}) \leq \beta(n)$;
- **No:** $\mathrm{T}(\rho, (I/2)^{\otimes n}) \geq \alpha(n)$.

The problem QSCMM provides a complete characterization of the class NIQSZK [Kob03, BST10, CCKV08], which consists of promise problems admitting non-interactive quantum statistical zero-knowledge proofs. It is widely believed that $\mathsf{BQP} \subsetneq \mathsf{NIQSZK}$. We next restate the corresponding NIQSZK-hardness result, which will later be used to establish the computational hardness of estimating the quantum q-Tsallis entropy for $0 < q < 1$ in Section 5.3:

Lemma 2.18 (QSCMM is NIQSZK-hard, adapted from [CCKV08, Section 8.1]). *For all integers $n \geq 3$, QSCMM$\left[\frac{1}{n}, 1 - \frac{1}{n}\right]$ is NIQSZK-hard.*

3 The Case of $q > 2$

3.1 Upper bound

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

[11] An $m(n)$-qubit maximally mixed state $(I/2)^{\otimes n}$ can be prepared by a quantum circuit that generates $m(n)$ EPR pairs on $m = 2n$ qubits and traces out one half of the qubits.

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 $\mathrm{tr}(\rho^q)$.

1: $\varepsilon' \leftarrow \varepsilon/(q+3)$, $m \leftarrow \min\{\lceil 1/\varepsilon' \rceil, d\}$, $\delta' \leftarrow 1/3m$.
2: $n \leftarrow \Theta(1/\varepsilon'^2)$, $k \leftarrow \Theta(\log(1/\delta'))$.
3: $\hat{\alpha} \leftarrow$ `SpectrumEstimation`(ρ, n, k).
4: $\hat{P} \leftarrow \sum_{j=1}^{m} \hat{\alpha}_j^q$.
5: **return** $\hat{P}$.

Theorem 3.1. *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.*

Proof. Our estimator for Theorem 3.1 is formally given in Algorithm 2, where we set

$$\varepsilon' := \frac{\varepsilon}{q+3}, \quad m = \min\left\{\left\lceil \frac{1}{\varepsilon'} \right\rceil, d\right\}, \quad \delta' := \frac{1}{3m}, \quad n = \Theta\left(\frac{1}{\varepsilon'^2}\right), \quad k = \Theta\left(\log\left(\frac{1}{\delta'}\right)\right). \tag{2}$$

By Lemma 2.14, we can use $nk = O(\log(1/\delta')/\varepsilon'^2)$ samples of ρ to obtain a sequence $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d)$ such that for every $j \in [d]$,

$$\Pr\left[|\hat{\alpha}_j - \alpha_j| \leq \varepsilon'\right] \geq 1 - \delta'. \tag{3}$$

Then, we consider the estimator:

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

The additive error is bounded by

$$\begin{aligned} \left|\hat{P} - \mathrm{tr}(\rho^q)\right| &= \left|\sum_{j=1}^{m}\left(\hat{\alpha}_j^q - \alpha_j^q\right) - \sum_{j=m+1}^{d} \alpha_j^q\right| \\ &\leq \sum_{j=1}^{m}\left|\hat{\alpha}_j^q - \alpha_j^q\right| + \sum_{j=m+1}^{d} \alpha_j^q. \end{aligned} \tag{4}$$

For the first term of Equation (4), note that

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

then we have

$$\begin{aligned} \left|\hat{\alpha}_j^q - \alpha_j^q\right| &\leq |\hat{\alpha}_j - \alpha_j||\hat{\alpha}_j|^{q-1} + |\alpha_j|\left|\hat{\alpha}_j^{q-1} - \alpha_j^{q-1}\right| \\ &\leq |\hat{\alpha}_j - \alpha_j||\hat{\alpha}_j| + |\alpha_j|(q-1)|\hat{\alpha}_j - \alpha_j| \qquad (5) \\ &\leq |\hat{\alpha}_j - \alpha_j|(|\alpha_j| + |\hat{\alpha}_j - \alpha_j|) + (q-1)|\alpha_j||\hat{\alpha}_j - \alpha_j| \\ &= q\alpha_j|\hat{\alpha}_j - \alpha_j| + |\hat{\alpha}_j - \alpha_j|^2, \qquad (6) \end{aligned}$$

where Equation (5) is by Fact 2.1. From Equation (6) and by Equation (3), the following holds with probability $\geq 1 - \delta'$:

$$\left|\hat{\alpha}_j^q - \alpha_j^q\right| \leq q\alpha_j\varepsilon' + \varepsilon'^2.$$

Therefore, we have that with probability $\geq 1 - m\delta'$, the following holds:

$$\begin{aligned}\sum_{j=1}^{m}\left|\hat{\alpha}_j^q - \alpha_j^q\right| &\leq \sum_{j=1}^{m}\left(q\alpha_j\varepsilon' + \varepsilon'^2\right) \\ &= q\varepsilon'\sum_{j=1}^{m}\alpha_j + m\varepsilon'^2 \\ &\leq q\varepsilon' + m\varepsilon'^2. \end{aligned} \tag{7}$$

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

$$\begin{aligned}\sum_{j=m+1}^{d}\alpha_j^q &\leq \sum_{j=m+1}^{d}\left(\frac{1}{j}\right)^q \\ &\leq \int_m^d \left(\frac{1}{x}\right)^q \mathrm{d}x \\ &= \frac{m^{1-q} - d^{1-q}}{q-1}. \end{aligned} \tag{8}$$

Combining Equations (7) and (8) in Equation (4), we have that with probability $\geq 1 - m\delta'$, the following holds:

$$\left|\hat{P} - \mathrm{tr}(\rho^q)\right| \leq q\varepsilon' + m\varepsilon'^2 + \frac{m^{1-q} - d^{1-q}}{q-1}. \tag{9}$$

According to the choice of ε', m, δ' in Equation (2) and by Equation (9), we have that with probability $\geq 1 - m\delta' = 2/3$, it holds that

$$\left|\hat{P} - \mathrm{tr}(\rho^q)\right| \leq \varepsilon.$$

To see this, we consider the following two cases:

1. $1/\varepsilon' \leq d$. In this case, $1/\varepsilon' \leq m = \lceil 1/\varepsilon' \rceil < 1/\varepsilon' + 1$. We have

$$\begin{aligned}(9) &\leq q\varepsilon' + \left(\frac{1}{\varepsilon'} + 1\right)\varepsilon'^2 + \frac{1}{m} \\ &\leq q\varepsilon' + \varepsilon' + \varepsilon'^2 + \varepsilon' \\ &\leq (q+3)\varepsilon' \\ &= \varepsilon. \end{aligned}$$

2. $1/\varepsilon' > d$. In this case, $m = d < 1/\varepsilon'$. We have

$$\begin{aligned}(9) &= q\varepsilon' + d\varepsilon'^2 \\ &\leq (q+1)\varepsilon' \\ &< \varepsilon. \end{aligned}$$

To complete the proof, the sample complexity is

$$O\left(\frac{\log(1/\delta')}{\varepsilon'^2}\right) = O\left(\frac{\log(1/\varepsilon)}{\varepsilon^2}\right).$$

□

3.2 Lower bound

Theorem 3.2. *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 2.5, we have $S = \Omega(1/\varepsilon^2)$. □

4 The Case of $1 < q < 2$

4.1 Upper bound

Algorithm 3 `PowerTrace`(ρ, q, ε) 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: $n \leftarrow \Theta(1/\varepsilon'^2)$, $k \leftarrow \Theta(\log(1/\delta'))$.
3: $\hat{\alpha} \leftarrow$ `SpectrumEstimation`(ρ, n, k).
4: $\hat{P} \leftarrow \sum_{j=1}^{m} \hat{\alpha}_j^q$.
5: **return** $\hat{P}$.

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

Theorem 4.1. *For every constant $1 < q < 2$, we can estimate $\mathrm{tr}(\rho^q)$ to within additive error ε using $O(\log(1/\varepsilon)/\varepsilon^{\frac{2}{q-1}})$ samples of ρ.*

To show Theorem 4.1, we need the following inequalities.

Lemma 4.2. *Suppose that* $1 < q < 2$ *and* $x_1 \geq x_2 \geq \cdots \geq x_N \geq 0$ *with* $\sum_{i=1}^N x_i = 1$. *For any positive integer* $m \leq N$, *we have*

$$\sum_{i=m+1}^{N} x_i^q \leq \frac{1}{m^{q-1}}.$$

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 $x_m \geq x_{m+1} \geq \cdots \geq x_N$, 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

$$\begin{aligned} \sum_{i=m+1}^{N} x_i^q &\leq l \cdot x_m^q + y^q \\ &= x_{m+1}^q \left(l + \left(\frac{A}{x_m} - l \right)^q \right) \\ &\leq x_m^q \cdot \frac{A}{x_m} \qquad (10) \\ &\leq x_m^{q-1}, \qquad (11) \end{aligned}$$

where Equation (10) is because $A/x_m - l < 1$. Then, by noting that $x_m \leq 1/m$, we have

$$(11) \leq \frac{1}{m^{q-1}}.$$

□

Now we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. Our estimator for Theorem 4.1 is formally given in Algorithm 3, where we set

$$\varepsilon' := \left(\frac{\varepsilon}{5}\right)^{\frac{1}{q-1}}, \quad m = \min\left\{\left\lceil \frac{1}{\varepsilon'} \right\rceil, d\right\}, \quad \delta' := \frac{1}{3m}, \quad n = \Theta\left(\frac{1}{\varepsilon'^2}\right), \quad k = \Theta\left(\log\left(\frac{1}{\delta'}\right)\right). \tag{12}$$

By Lemma 2.14, we can use $nk = O(\log(1/\delta')/\varepsilon'^2)$ samples of ρ to obtain a sequence $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_d)$ such that for every $j \in [d]$,

$$\Pr\left[|\hat{\alpha}_j - \alpha_j| \le \varepsilon'\right] \ge 1 - \delta'. \tag{13}$$

Then, we consider the estimator:

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

We have

$$\begin{aligned} \left|\hat{P} - \operatorname{tr}(\rho^q)\right| &= \left|\sum_{j=1}^{m} \left(\hat{\alpha}_j^q - \alpha_j^q\right) - \sum_{j=m+1}^{d} \alpha_j^q\right| \\ &\le \sum_{j=1}^{m} \left|\hat{\alpha}_j^q - \alpha_j^q\right| + \sum_{j=m+1}^{d} \alpha_j^q. \end{aligned} \tag{14}$$

For the first term of Equation (14), note that

$$\begin{aligned} \left|\hat{\alpha}_j^q - \alpha_j^q\right| &= \left|(\hat{\alpha}_j - \alpha_j)\hat{\alpha}_j^{q-1} + \alpha_j(\hat{\alpha}_j^{q-1} - \alpha_j^{q-1})\right| \\ &\le |\hat{\alpha}_j - \alpha_j|\hat{\alpha}_j^{q-1} + \alpha_j\left|\hat{\alpha}_j^{q-1} - \alpha_j^{q-1}\right| \\ &\le |\hat{\alpha}_j - \alpha_j|\hat{\alpha}_j^{q-1} + \alpha_j|\hat{\alpha}_j - \alpha_j|^{q-1}, \end{aligned} \tag{15}$$

where the last inequality is by Fact 2.2. Then, by Equation (13), with probability $\ge 1 - \delta'$, the following holds:

$$(15) \le \varepsilon'(\alpha_j + \varepsilon')^{q-1} + \alpha_j(\varepsilon')^{q-1}.$$

This implies, with probability $\ge 1 - m\delta'$, we have

$$\begin{aligned} \sum_{j=1}^{m}\left|\hat{\alpha}_j^q - \alpha_j^q\right| &\le \varepsilon' \sum_{j=1}^{m}(\alpha_j + \varepsilon')^{q-1} + (\varepsilon')^{q-1}\sum_{j=1}^{m}\alpha_j \\ &\le \varepsilon' \sum_{j=1}^{m}(\alpha_j + \varepsilon')^{q-1} + (\varepsilon')^{q-1} \\ &\le \varepsilon' m^{2-q} \cdot \left(m\varepsilon' + \sum_{j=1}^{m}\alpha_j\right)^{q-1} + (\varepsilon')^{q-1} && (16) \\ &\le \varepsilon' m^{2-q}\left(m\varepsilon' + 1\right)^{q-1} + (\varepsilon')^{q-1}, && (17) \end{aligned}$$

where Equation (16) is by Fact 2.3.

Combining Equation (17) with Equation (14), we have that, with probability $\geq 1 - m\delta'$, it holds that

$$\left|\hat{P} - \operatorname{tr}(\rho^q)\right| \leq \varepsilon' m^{2-q}\big(m\varepsilon' + 1\big)^{q-1} + (\varepsilon')^{q-1} + \sum_{j=m+1}^{d} \alpha_j^q. \tag{18}$$

According to the choice of ε', m, δ' in Equation (12) and by Equation (18), we have 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 4.2 to obtain:

$$\sum_{j=m+1}^{d} \alpha_j^q \leq \frac{1}{m^{q-1}}. \tag{19}$$

   Using Equation (19), we have

$$\begin{aligned}
(18) &\leq \varepsilon'(2/\varepsilon')^{2-q}(2+1)^{q-1} + (\varepsilon')^{q-1} + (1/\varepsilon')^{1-q} \\
&\leq 3(\varepsilon')^{q-1} + 2(\varepsilon')^{q-1} \\
&\leq 5(\varepsilon')^{q-1} \\
&= \varepsilon.
\end{aligned}$$

2. $1/\varepsilon' > d$. In this case, $m = d < 1/\varepsilon'$. We have

$$\begin{aligned}
(18) &\leq \varepsilon'(1/\varepsilon')^{2-q}(1+1)^{q-1} + (\varepsilon')^{q-1} \\
&\leq 5(\varepsilon')^{q-1} \\
&= \varepsilon.
\end{aligned}$$

To complete the proof, the sample complexity is

$$O\left(\frac{\log(1/\delta')}{\varepsilon'^2}\right) = O\left(\frac{\log(1/\varepsilon)}{\varepsilon^{\frac{2}{q-1}}}\right).$$

□

4.2 Lower bound

Theorem 4.3. *For every constant $1 < q < 2$, when the dimension of ρ is sufficiently large, any quantum estimator to additive error ε for $\operatorname{tr}(\rho^q)$ requires sample complexity $\Omega(1/\varepsilon^{\frac{1}{q-1}})$.*

Proof. Given integers $1 \leq r \leq d$, we use $D_{r,d}$ to denote the $d \times d$ diagonal matrix:

$$D_{r,d} \coloneqq \operatorname{diag}(\underbrace{\tfrac{1}{r}, \ldots, \tfrac{1}{r}}_{r}, \underbrace{0, \ldots, 0}_{d-r}).$$

Let

$$r = \left\lfloor \frac{1}{(2\varepsilon)^{\frac{1}{q-1}}} \right\rfloor \quad \text{and} \quad d = \left\lfloor \frac{1}{\varepsilon^{\frac{1}{q-1}}} \right\rfloor + 1.$$

If the number of samples $n > r$, then we directly have $n \geq \Omega(1/\varepsilon^{\frac{1}{q-1}})$. Therefore, we assume

$$n \leq r. \tag{20}$$

Then, consider the following distinguishing problem.

Problem 1. *Suppose a d-dimensional quantum state ρ is in one of the following with equal probability:*

1. $\rho = \rho_1 \coloneqq U D_{r,d} U^\dagger$, *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

$$\begin{aligned}
\mathbb{E}\big[\rho_1^{\otimes n}\big] &= \mathop{\mathbb{E}}_{U \sim \mathbb{U}_d}\Big[U^{\otimes n} D_{r,d}^{\otimes n} U^{\dagger \otimes n}\Big] \\
&= \bigoplus_{\lambda \vdash n} I_{\mathcal{P}_\lambda} \otimes \mathop{\mathbb{E}}_{U \sim \mathbb{U}_d}\Big[\mathsf{q}_\lambda(U)\mathsf{q}_\lambda(D_{r,d})\mathsf{q}_\lambda(U)^\dagger\Big] &(21)\\
&= \bigoplus_{\lambda \vdash n} I_{\mathcal{P}_\lambda} \otimes I_{\mathcal{Q}_\lambda^d} \cdot \frac{s_\lambda(D_{r,d})}{\dim(\mathcal{Q}_\lambda^d)}, &(22)
\end{aligned}$$

where Equation (21) can be seen by Fact 2.9, in Equation (22) is by Corollary 2.7 and the observation that $\mathbb{E}_{U \sim \mathbb{U}_d}\big[\mathsf{q}_\lambda(U)\mathsf{q}_\lambda(D_{r,d})\mathsf{q}_\lambda(U)^\dagger\big]$ 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

$$\mathbb{E}\big[\rho_2^{\otimes n}\big] = \bigoplus_{\lambda \vdash n} I_{\mathcal{P}_\lambda} \otimes I_{\mathcal{Q}_\lambda^d} \cdot \frac{s_\lambda(D_{d,d})}{\dim(\mathcal{Q}_\lambda^d)}.$$

By Theorem 2.4, the success probability of distinguishing $\mathbb{E}[\rho_1^{\otimes n}]$ and $\mathbb{E}[\rho_2^{\otimes n}]$ is upper bounded by

$$\frac{1}{2} + \frac{1}{4}\big\|\mathbb{E}\big[\rho_1^{\otimes n}\big] - \mathbb{E}\big[\rho_2^{\otimes n}\big]\big\|_1.$$

Note that

$$\begin{aligned}
\big\|\mathbb{E}\big[\rho_1^{\otimes n}\big] - \mathbb{E}\big[\rho_2^{\otimes n}\big]\big\|_1 &= \sum_{\lambda \vdash n} |\dim(\mathcal{P}_\lambda) \cdot s_\lambda(D_{r,d}) - \dim(\mathcal{P}_\lambda) \cdot s_\lambda(D_{d,d})| \\
&= \| \operatorname{SW}_r^n - \operatorname{SW}_d^n \|_1, &(23)
\end{aligned}$$

where in Equation (23) we use the stability of Schur polynomial, i.e.,

$$s_\lambda(D_{r,d}) = s_\lambda(\underbrace{\tfrac{1}{r}, \ldots, \tfrac{1}{r}}_{r}, \underbrace{0, \ldots, 0}_{d-r}) = s_\lambda(\underbrace{\tfrac{1}{r}, \ldots, \tfrac{1}{r}}_{r}).$$

Then, since $n \leq r \leq d$ (see Equation (20)), by Lemma 2.13, we have that

$$\| \operatorname{SW}_r^n - \operatorname{Planch}(n)\|_1 \leq \sqrt{2}\frac{n}{r},$$

and

$$\|\operatorname{SW}_d^n - \operatorname{Planch}(n)\|_1 \le \sqrt{2}\frac{n}{d}.$$

This means

$$\begin{aligned}\|\operatorname{SW}_r^n - \operatorname{SW}_d^n\|_1 &\le \|\operatorname{SW}_r^n - \operatorname{Planch}(n)\|_1 + \|\operatorname{SW}_d^n - \operatorname{Planch}(n)\|_1 \\ &\le \sqrt{2}\frac{n}{r} + \sqrt{2}\frac{n}{d}.\end{aligned}$$

Therefore, if the success probability is at least 2/3, we have

$$\frac{2}{3} \le \frac{1}{2} + \frac{1}{4}\left(\sqrt{2}\frac{n}{r} + \sqrt{2}\frac{n}{d}\right) \le \frac{1}{2} + \frac{n}{\sqrt{2}r},$$

which means

$$n \ge \frac{\sqrt{2}r}{6} = \frac{\sqrt{2}}{6}\left\lfloor \frac{1}{(2\varepsilon)^{\frac{1}{q-1}}} \right\rfloor = \Omega\left(\frac{1}{\varepsilon^{\frac{1}{q-1}}}\right).$$

□

5 The Case of $0 < q < 1$

5.1 Upper bound

Algorithm 4 `PowerTrace`(ρ, q, ε) for $0 < q < 1$

Input: Sample access to a d-dimensional mixed quantum state ρ; $q \in (0,1)$ and $\varepsilon \in (0,1)$.
Output: An estimate of $\operatorname{tr}(\rho^q)$.
1: $\eta \leftarrow \varepsilon^{1/q}/d^{1/q-1}$, $n \leftarrow \Theta(d^2/\eta^2)$.
2: $\hat{\alpha} \leftarrow$ `SpectrumEstimation`$(\rho, n, 1)$.
3: $\hat{P} \leftarrow \sum_{j=1}^{d} \hat{\alpha}_j^q$.
4: **return** $\hat{P}$.

We state the sample complexity of estimating $\operatorname{tr}(\rho^q)$ for the case of $0 < q < 1$ as follows.

Theorem 5.1. *For every constant $0 < q < 1$, we can estimate $\operatorname{tr}(\rho^q)$ to within additive error ε using $O(d^{2/q}/\varepsilon^{2/q})$ samples of ρ.*

Proof. Let $\eta := \varepsilon^{1/q}/d^{1/q-1}$. By Lemma 2.15, we can use $n := \Theta(d^2/\eta^2) = O((d/\varepsilon)^{2/q})$ samples of ρ to obtain a sequence $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_d)$ such that

$$\sum_{j=1}^{d} |\hat{\alpha}_j - \alpha_j| \le \eta,$$

with probability at least 2/3. Then, we consider the estimator

$$\hat{P} := \sum_{j=1}^{d} \hat{\alpha}_j^q.$$

We have

$$\begin{aligned}
\left|\hat{P} - \operatorname{tr}(\rho^q)\right| &\le \sum_{j=1}^{d} \left|\hat{\alpha}_j^q - \alpha_j^q\right| \\
&\le \sum_{j=1}^{d} |\hat{\alpha}_j - \alpha_j|^q && (24) \\
&\le \left(\sum_{j=1}^{d} |\hat{\alpha}_j - \alpha_j|\right)^q d^{1-q} && (25) \\
&\le \eta^q d^{1-q} = \varepsilon,
\end{aligned}$$

where Equation (24) is due to Fact 2.2, and Equation (25) is due to Fact 2.3. Therefore, with probability at least $2/3$, $\hat{P}$ gives a good estimate for $\operatorname{tr}(\rho^q)$. □

5.2 Lower bound

Theorem 5.2. *For every constant $0 < q < 1$ and sufficiently large d, any quantum estimator to additive error ε for $\operatorname{tr}(\rho^q)$ requires sample complexity $\Omega(d^{1/q}/\varepsilon^{1/q})$.*

Before giving the proof, we first introduce the following notations. Let n, m, d be positive integers such that $n \ge m$. Let $\mathcal{H}_1 \cong \cdots \cong \mathcal{H}_n \cong \mathbb{C}^d$ be n copies of the d-dimensional Hilbert space. Let $S \subseteq [n] = \{1, 2, \ldots, n\}$ be a set of integers and A be a $d \times d$ matrix. We use the following notation $A^{\otimes S}$ to denote the matrix $A^{\otimes |S|}$ acting on $\bigotimes_{i \in S} \mathcal{H}_i$. Therefore, if B is another $d \times d$ matrix, then $A^{\otimes S} \otimes B^{\otimes [n] \setminus S}$ denotes the matrix $\bigotimes_{i=1}^n X_i$ acting on $\bigotimes_{i=1}^n \mathcal{H}_i$ where $X_i = A$ for $i \in S$, and $X_i = B$ otherwise. Now we give the proof below.

Proof of Theorem 5.2. Given an integer $r \in [1, d-1]$, and $\Delta \in [0, 1]$, we use D_r to denote the $(d-1) \times (d-1)$ diagonal matrix:

$$D_r \coloneqq \operatorname{diag}(\underbrace{\tfrac{1}{r}, \ldots, \tfrac{1}{r}}_{r}, \underbrace{0, \ldots, 0}_{d-1-r}),$$

with respect to the orthonormal states $\{|1\rangle, \ldots, |d-1\rangle\}$ (i.e., $D_r = \frac{1}{r}\sum_{i=1}^r |i\rangle\langle i|$). Let

$$r_1 \coloneqq \left\lfloor \frac{d-1}{2} \right\rfloor, \quad r_2 \coloneqq 2r_1, \quad \text{and} \quad \Delta \coloneqq \frac{\varepsilon^{1/q}}{d^{1/q-1}} \cdot \frac{1}{(1 - \frac{1}{2^{1-q}})^{1/q}} \le \frac{\varepsilon^{1/q}}{d^{1/q-1}} \cdot O\left(\frac{1}{1-q}\right).$$

Then, consider the following distinguishing problem.

Problem 2. *Suppose a d-dimensional quantum state ρ is in one of the following with equal probability:*

1. $\rho = \rho_1 \coloneqq (1 - \Delta) \cdot |0\rangle\langle 0| + \Delta \cdot U D_{r_1} U^\dagger$,
2. $\rho = \rho_2 \coloneqq (1 - \Delta) \cdot |0\rangle\langle 0| + \Delta \cdot U D_{r_2} U^\dagger$,

where $U \sim \mathbb{U}_{d-1}$ is a $d-1$-dimensional Haar random unitary acting on $\operatorname{span}\{|1\rangle, \ldots, |d-1\rangle\}$. *The task is to distinguish between the above two cases.*

Suppose an algorithm can estimate $\mathrm{tr}(\rho^q)$ to within error $\varepsilon/10$, with sample complexity n. Then, this algorithm can distinguish these two cases, since

$$\begin{aligned}\mathrm{tr}(\rho_2^q) - \mathrm{tr}(\rho_1^q) &= \Delta^q r_2^{1-q} - \Delta^q r_1^{1-q} \\ &= \frac{\varepsilon}{d^{1-q}(1-\frac{1}{2^{1-q}})}(r_2^{1-q} - r_1^{1-q}) \\ &= \frac{\varepsilon r_2^{1-q}}{d^{1-q}} \\ &\geq \frac{\varepsilon}{2}.\end{aligned}$$

However, to distinguish these two states, it must satisfy that:

$$n \geq \Omega(d/\Delta) = \Omega\left((1-q)\frac{d^{1/q}}{\varepsilon^{1/q}}\right).$$

To see this, we first assume that $10000 n\Delta \leq d$, since otherwise $n \geq \Omega(d/\Delta)$ and we have done. Note that

$$\begin{aligned}\mathop{\mathbb{E}}_{U\sim\mathbb{U}_{d-1}}[\rho_1^{\otimes n}] &= \mathop{\mathbb{E}}_{U\sim\mathbb{U}_{d-1}}\left[\sum_{i=0}^{n}\Delta^i(1-\Delta)^{n-i}\cdot\sum_{\substack{S\subseteq[n]\\|S|=i}}\left(UD_{r_1}U^\dagger\right)^{\otimes S}\otimes|0\rangle\langle 0|^{\otimes[n]\setminus S}\right] \\ &= \sum_{i=0}^{n}\Delta^i(1-\Delta)^{n-i}\cdot\sum_{\substack{S\subseteq[n]\\|S|=i}}\overline{D}_{r_1}^S\otimes|0\rangle\langle 0|^{\otimes[n]\setminus S},\end{aligned}$$

where $\overline{D}_r^S \coloneqq \mathbb{E}_{U\sim\mathbb{U}_{d-1}}\left[(UD_rU^\dagger)^{\otimes S}\right]$. Similarly,

$$\mathop{\mathbb{E}}_{U\sim\mathbb{U}_{d-1}}[\rho_2^{\otimes n}] = \sum_{i=0}^{n}\Delta^i(1-\Delta)^{n-i}\cdot\sum_{\substack{S\subseteq[n]\\|S|=i}}\overline{D}_{r_2}^S\otimes|0\rangle\langle 0|^{\otimes[n]\setminus S}.$$

Then, we can see that

$$\begin{aligned}\left\|\mathop{\mathbb{E}}_{U}[\rho_1^{\otimes n}] - \mathop{\mathbb{E}}_{U}[\rho_2^{\otimes n}]\right\|_1 &\leq \sum_{i=0}^{n}\Delta^i(1-\Delta)^{n-i}\cdot\sum_{\substack{S\subseteq[n]\\|S|=i}}\left\|\overline{D}_{r_1}^S - \overline{D}_{r_2}^S\right\|_1 \\ &= \sum_{i=1}^{n}\Delta^i(1-\Delta)^{n-i}\cdot\binom{n}{i}\cdot\left\|\overline{D}_{r_1}^i - \overline{D}_{r_2}^i\right\|_1 &(26)\\ &\leq \sum_{1\leq i\leq\min\{n,100n\Delta\}}\Delta^i(1-\Delta)^{n-i}\cdot\binom{n}{i}\cdot\left\|\overline{D}_{r_1}^i - \overline{D}_{r_2}^i\right\|_1 + \frac{e}{100} &(27)\\ &= \sum_{1\leq i\leq\min\{n,100n\Delta\}}\Delta^i(1-\Delta)^{n-i}\cdot\binom{n}{i}\cdot\left\|\mathrm{SW}_{r_1}^i - \mathrm{SW}_{r_2}^i\right\|_1 + \frac{e}{100} &(28)\\ &\leq \sum_{1\leq i\leq\min\{n,100n\Delta\}}\Delta^i(1-\Delta)^{n-i}\cdot\binom{n}{i}\cdot\left(\sqrt{2}\frac{i}{r_1} + \sqrt{2}\frac{i}{r_2}\right) + \frac{e}{100} &(29)\end{aligned}$$

$$\leq \frac{1000n\Delta}{d} + \frac{e}{100}, \tag{30}$$

where in Equation (26) we define $\overline{D}_r^i = \mathbb{E}_{U\sim\mathbb{U}_{d-1}}\left[(UD_rU^\dagger)^{\otimes i}\right]$, Equation (27) is due to Lemma 5.3, Equation (28) is by noting that (similar to Equation (21) and Equation (22))

$$\overline{D}_r^i = \mathop{\mathbb{E}}_{U\sim\mathbb{U}_{d-1}}\left[U^{\otimes i}D_r^{\otimes i}U^{\dagger\otimes i}\right] = \bigoplus_{\lambda\vdash n} I_{\mathcal{P}_\lambda}\otimes I_{\mathcal{Q}_\lambda^{d-1}}\cdot\frac{s_\lambda(D_r)}{\dim(\mathcal{Q}_\lambda^{d-1})},$$

Equation (29) is because

$$\left\|\mathrm{SW}_{r_1}^i - \mathrm{SW}_{r_2}^i\right\| \leq \left\|\mathrm{SW}_{r_1}^i - \mathrm{Planch}(i)\right\| + \left\|\mathrm{SW}_{r_2}^i - \mathrm{Planch}(i)\right\| \leq \sqrt{2}\frac{i}{r_1} + \sqrt{2}\frac{i}{r_2},$$

due to Lemma 2.13 and the fact that $i \leq 100n\Delta \leq d/100 < r_1 < r_2$, and Equation (30) is because $\sum_{1\leq i\leq\min\{n,100n\Delta\}}\Delta^i(1-\Delta)^{n-i}\binom{n}{i} \leq 1$. Therefore, if the success probability is at least 2/3, then due to Theorem 2.4, we have

$$\frac{2}{3} \leq \frac{1}{2} + \frac{1}{4}\left(\frac{1000n\Delta}{d} + \frac{e}{100}\right),$$

which means

$$n \geq \Omega(d/\Delta),$$

as desired. □

Lemma 5.3. *Let $n > 0$ be an integer and $p \in [0,1]$. Then, for any $k > e$, we have*

$$\sum_{knp<i\leq n} p^i(1-p)^{n-i}\binom{n}{i} \leq \frac{e}{k}.$$

Proof. Let $X_1,\ldots,X_n$ be independent Bernoulli random variables such that $\Pr[X_i = 1] = p$. Let $X = \sum_{i=1}^n X_i$ and note that $\mathbb{E}[X] = np$. If $p = 0$, then the claim is trivial. Hence, we may assume $p > 0$. Let $m = \max\{1, knp\}$ and $\gamma = m/np$. Then, we have

$$\begin{aligned}\sum_{knp<i\leq n} p^i(1-p)^{n-i}\binom{n}{i} &\leq \sum_{m\leq i\leq n} p^i(1-p)^{n-i}\binom{n}{i} \leq \Pr[X\geq m]\\ &\overset{\text{(i)}}{\leq} \left(\frac{e^{\gamma-1}}{\gamma^\gamma}\right)^{np} \leq \left(\frac{e}{\gamma}\right)^{\gamma np}\\ &\overset{\text{(ii)}}{\leq} \frac{e}{k},\end{aligned}$$

where (i) is by the Chernoff's bound, and (ii) is because $\gamma \geq k > e$ and $\gamma np \geq 1$. □

5.3 Computational Hardness

We begin with the promise problems corresponding to estimating the quantum q-Tsallis entropy:

Definition 5.4 (Quantum q-Tsallis Entropy Approximation, $\textsc{TsallisQEA}_q$, adapted from [LW25, Definition 5.2])**.** *Let Q be a quantum circuit acting on m qubits with n specified output qubits, where $m \geq 2n$. Let ρ denote the quantum state obtained by applying Q to the initial state $|0\rangle^{\otimes m}$ and tracing out the non-output qubits. Let $g(n)$ and $t(n)$ be positive efficiently computable functions. The task is to decide whether the following holds:*

- **Yes:** $\mathrm{S}_q^{\mathrm{T}}(\rho) \geq t(n) + g(n)$;
- **No:** $\mathrm{S}_q^{\mathrm{T}}(\rho) \leq t(n) - g(n)$.

The main result of this subsection is the following hardness statement, which directly follows from a reduction from QSCMM to $\textsc{TsallisQEA}_q$ for $q \in (0,1)$, as detailed in Lemma 5.11:

Theorem 5.5 ($\textsc{TsallisQEA}_q$ is NIQSZK-hard for $0 < q < 1$). *For all real $q \in (0,1)$ and all integers $n \geq \lceil \frac{5}{q(1-q)} \rceil$, the following holds:*

$$\forall g(n;q) \in \left[\frac{1}{\mathrm{poly}(n)}, \frac{2^{5/q} q}{40}\right], \ \textsc{TsallisQEA}_q[t(n;q), g(n;q)] \text{ is } \mathsf{NIQSZK}\text{-hard.}$$

The threshold parameter $t(n;q)$ is as defined in Equation (41).

To establish the reduction QSCMM $\leq \textsc{TsallisQEA}_q$, we prove an inequality that relates the q-Tsallis entropy of a distribution to its total variation distance from the uniform distribution. Specifically, combining Lemmas 5.7 and 5.8 yields the desired bounds:[12]

Theorem 5.6 (Bounds on q-Tsallis entropy via closeness to uniform distribution for $0 < q < 1$). *Let p be a probability distribution over $[N]$, where $N \geq 2$, and let ν denote the uniform distribution over $[N]$. Then, for every $q \in (0,1)$, the following inequalities hold:*

$$\left(1 - \mathrm{TV}(p,\nu) - \frac{1}{N}\right) \ln_q(N) \leq \mathrm{H}_q^{\mathrm{T}}(p) \leq \ln_q(N) - \frac{q}{2} N^{1-q} \cdot \mathrm{TV}(p,\nu)^2.$$

In the remainder of this subsection, we present the proofs of Theorems 5.5 and 5.6. Notably, for $q = 1/2$, the upper bound in Theorem 5.6 can be strengthened using a simpler argument (Lemma 5.9). Consequently, this yields a slightly simpler reduction QSCMM $\leq \textsc{TsallisQEA}_{1/2}$ (Lemma 5.10) and the corresponding hardness result (Theorem 5.12).

5.3.1 Useful bounds on Tsallis entropy

In this subsection, we establish bounds on the q-Tsallis entropy, as stated in Lemmas 5.7 to 5.9.

Lemma 5.7 (Lower bound on q-Tsallis entropy via closeness to uniform distribution for $0 < q < 1$). *Let p be a probability distribution over $[N]$, where $N \geq 2$, and let ν denote the uniform distribution over $[N]$. Then, for every $q \in (0,1)$, the following inequality holds:*

$$\left(1 - \mathrm{TV}(p,\nu) - \frac{1}{N}\right) \ln_q(N) \leq \mathrm{H}_q^{\mathrm{T}}(p).$$

Proof. Let Δ_N denote the set of probability distributions over N elements. For any distribution $p \in \Delta_N$, it is straightforward to verify that $0 \leq \mathrm{TV}(p,\nu) \leq 1 - 1/N$. Throughout, we use the shorthand $\gamma \coloneqq \mathrm{TV}(p,\nu)$ for convenience.

To demonstrate the lower bound, we recall that $\mathrm{H}_q^{\mathrm{T}}(p)$ is concave for all $0 < q < 1$ (see, e.g., [LW25, Lemma 2.3]). Consequently, the proof strategy used for the cases $q = 1$ and $q > 1$, particularly those in [KLN19, Lemma 16] and [LW25, Lemma 4.10], applies in the present setting as well. Specifically, it suffices to solve the convex optimization problem in Equation (31), whose objective is to minimize the Tsallis entropy $\mathrm{H}_q^{\mathrm{T}}(p)$ under the constraint $\mathrm{TV}(p,\nu) = \gamma$. Since the

[12] Related results have previously been shown: for the case $q = 1$, see in [KLN19, Lemma 16]; and for the case $q > 1$ with the additional condition $\mathrm{TV}(p,\nu) \geq 1/q$ (tailored to the upper bound), see [LW25, Lemma 4.10].

feasible region specified in Equation (31) constitutes a closed convex set, the problem admits an optimal solution, one such minimizer $p_{\min}$ is given in Equation (32).[13]

$$\begin{aligned} &\text{minimize} && \mathrm{H}_q^{\mathrm{T}}(p') \\ &\text{subject to} && p' \in \Delta_N, \\ & && \mathrm{TV}(p',\nu) \le \gamma \end{aligned} \tag{31}$$

$$p_{\min}(i) = \begin{cases} \frac{1}{N}, & \text{if } i \in [k_{\min}] \\ \frac{1}{N}+\gamma, & \text{if } i = k_{\min}+1 \\ \frac{\varepsilon}{N}, & \text{if } i = k_{\min}+2 \\ 0, & \text{otherwise} \end{cases}, \tag{32a}$$

$$\text{where} \quad k_{\min} := \lfloor N(1-\gamma) \rfloor - 1, \quad \varepsilon := N(1-\gamma) - \lfloor N(1-\gamma) \rfloor. \tag{32b}$$

Next, we derive the lower bound for the Tsallis entropy by evaluating $\mathrm{H}_q^{\mathrm{T}}(p_{\min})$:

$$\begin{aligned} \mathrm{H}_q^{\mathrm{T}}(p_{\min}) &= \frac{1}{1-q}\left(k_{\min}\left(\frac{1}{N}\right)^q + \left(\frac{1}{N}+\gamma\right)^q + \left(\frac{\varepsilon}{N}\right)^q - 1\right) \\ &\ge \frac{1}{1-q}\left((\lfloor N(1-\gamma)\rfloor - 1 + \varepsilon^q)N^{-q} - \frac{(\lfloor N(1-\gamma)\rfloor - 1)}{N} - \frac{\varepsilon}{N}\right) \\ &= \frac{1}{1-q}\left((N(1-\gamma) - \varepsilon - 1 + \varepsilon^q)N^{-q} - \frac{N(1-\gamma)-1}{N}\right) \\ &\ge \frac{1}{1-q}\left(\left(1-\gamma-\frac{1}{N}\right)N^{1-q} - \left(1-\gamma-\frac{1}{N}\right)\right) \\ &= \left(1-\gamma-\frac{1}{N}\right)\ln_q(N). \end{aligned}$$

Here, in the second line, we omit the terms corresponding to $p_{\min}(k_{\min}+1)$, since $\left(\frac{1}{N}+\gamma\right)^q - \left(\frac{1}{N}+\gamma\right) \ge 0$ for $0 < q < 1$. The third line follows from the identity $\lfloor N(1-\gamma)\rfloor = N(1-\gamma) - \varepsilon$. The fourth line holds because $\varepsilon^q - \varepsilon \ge 0$ for $0 < q < 1$ and $0 \le \varepsilon \le 1$. □

Lemma 5.8 (Upper bound on q-Tsallis entropy via closeness to uniform distribution for $0 < q < 1$)**.** *Let p be a probability distribution over $[N]$, where $N \ge 2$, and let ν denote the uniform distribution over $[N]$. Then, for every $q \in (0,1)$, the following inequality holds:*

$$\mathrm{H}_q^{\mathrm{T}}(p) \le \ln_q(N) - \frac{q}{2}N^{1-q} \cdot \mathrm{TV}(p,\nu)^2.$$

Proof. As in Lemma 5.7, for any distribution $p \in \Delta_N$, we have $0 \le \gamma := \mathrm{TV}(p,\nu) \le 1 - 1/N$. To prove the upper bound, it therefore suffices to maximize the Tsallis entropy $\mathrm{H}_q^{\mathrm{T}}(p)$ under the constraint $\mathrm{TV}(p,\nu) = \gamma$, which can be formulated as a *non-convex* optimization problem analogous to Equation (31). Accordingly, we consider the following optimization problem:

$$\begin{aligned} &\text{maximize} && \mathrm{H}_q^{\mathrm{T}}(p') \\ &\text{subject to} && p' \in \Delta_N, \\ & && \mathrm{TV}(p',\nu) \ge \gamma. \end{aligned} \tag{33}$$

Although Equation (33) is analyzed in [LW25, Lemma 4.10] for the regime $q > 1$, the case $0 < q < 1$ requires a different analysis. Since $\mathrm{H}_q^{\mathrm{T}}(p)$ is maximized at ν, any optimizer of Equation (33)

[13] Since $\mathrm{H}_q^{\mathrm{T}}(p)$ is concave, its minimum over the compact convex polytope in Equation (31) is attained at an extreme point. A direct check of extremality shows that, up to permutation, the extreme points are exactly those of the form given in Equation (32).

may be taken to satisfy $\mathrm{TV}(p_{\max}, \nu) = \gamma$, and by Jensen's inequality, the optimizer must be equalized on each side of $1/N$. Hence, any optimal solution $p_{\max}$ must have the form:[14]

$$\mathrm{H}_q^{\mathrm{T}}(p_{\max}) = \max_{k \in [\lfloor N(1-\gamma) \rfloor]} \mathrm{H}_q^{\mathrm{T}}\left(p^{(k)}\right), \text{ where } p^{(k)}(i) := \begin{cases} \frac{1}{N} + \frac{\gamma}{k}, & \text{if } i \in [k], \\ \frac{1}{N} - \frac{\gamma}{N-k}, & \text{otherwise.} \end{cases} \tag{34}$$

Combining Equations (33) and (34), the task reduces to the following optimization problem for $0 < q < 1$, where the optimal choice of k depends on the order q:

$$\begin{aligned} \underset{k}{\text{maximize}} \quad & \mathrm{PS}_q(N, k, \gamma) := \sum_{i \in [N]} p^{(k)}(i)^q = k \cdot \left(\frac{1}{N} + \frac{\gamma}{k}\right)^q + (N-k) \cdot \left(\frac{1}{N} - \frac{\gamma}{N-k}\right)^q \\ \text{subject to} \quad & 0 \le \gamma \le 1 - 1/N, \\ & 1 \le k \le \lfloor N(1-\gamma) \rfloor, \\ & k, N \in \mathbb{Z}_+ \end{aligned} \tag{35}$$

In the remainder of the proof, we derive an upper bound on the optimal value of Equation (35). To this end, set $t := k/N \in [1/N, 1-\gamma]$. A straightforward calculation shows that

$$\begin{aligned} N^{q-1} \cdot \mathrm{PS}_q(N, k, \gamma) &= \left(\frac{k}{N}\right)^{1-q} \left(\frac{k}{N} + \gamma\right)^q + \left(\frac{N-k}{N}\right)^{1-q} \left(\frac{N-k}{N} - \gamma\right)^q \\ &= t^{1-q}(t+\gamma)^q + (1-t)^{1-q}(1-t-\gamma)^q := F_q(t, \gamma). \end{aligned}$$

Applying Taylor's theorem to $F_q(t, \gamma)$ around $\gamma = 0$, there exists $\xi \in (0, \gamma)$ corresponding to the Lagrange remainder such that

$$F_q(t, \gamma) = F_q(t, 0) + \gamma \cdot \frac{\partial}{\partial \gamma} F_q(t, \gamma)\bigg|_{\gamma=0} + \frac{\gamma^2}{2} \cdot \frac{\partial^2}{\partial \gamma^2} F_q(t, \gamma)\bigg|_{\gamma=\xi}. \tag{36}$$

The first and second derivatives are given by

$$\frac{\partial}{\partial \gamma} F_q(t, \gamma) = q t^{1-q}(t+\gamma)^{q-1} - q(1-t)^{1-q}(1-t-\gamma)^{q-1}, \tag{37a}$$

$$\frac{\partial^2}{\partial \gamma^2} F_q(t, \gamma) = q(q-1)\big(t^{1-q}(t+\gamma)^{q-2} + (1-t)^{1-q}(1-t-\gamma)^{q-2}\big). \tag{37b}$$

Combining Equations (36) and (37), to establish the desired inequality it suffices to prove that

$$1 + \frac{\gamma^2}{2} \cdot \frac{\partial^2}{\partial \gamma^2} F_q(t, \gamma)\bigg|_{\gamma=\xi} \le 1 - \frac{q}{2}(1-q)\gamma^2. \tag{38}$$

We now establish Equation (38) by deriving the following lower bound via direct calculation:

$$\frac{1}{q(q-1)} \cdot \frac{\partial^2}{\partial \gamma^2} F_q(t, \gamma)\bigg|_{\gamma=\xi} = t^{1-q}(t+\xi)^{q-2} + (1-t)^{1-q}(1-t-\xi)^{q-2} \tag{39a}$$

---

[14] To see this, let $p' \in \Delta_N$ be feasible for Equation (33), and write $\gamma' := \mathrm{TV}(p', \nu) \ge \gamma$. If $\gamma' > \gamma$, then for $\tilde{p} := \frac{\gamma}{\gamma'} p' + \big(1 - \frac{\gamma}{\gamma'}\big)\nu$, we have $\mathrm{TV}(\tilde{p}, \nu) = \gamma$, and by concavity of $\mathrm{H}_q^{\mathrm{T}}(p)$ together with the fact that $\mathrm{H}_q^{\mathrm{T}}(\nu)$ is maximal, $\mathrm{H}_q^{\mathrm{T}}(\tilde{p}) \ge \mathrm{H}_q^{\mathrm{T}}(p')$. Hence it suffices to consider the case $\mathrm{TV}(p', \nu) = \gamma$. Now let $A := \{i : p'(i) > 1/N\}$ and $k := |A|$. Then $\sum_{i \in A}(p'(i) - 1/N) = \gamma$ and $\sum_{i \notin A}(1/N - p'(i)) = \gamma$. Since $x \mapsto x^q$ is concave for $0 < q < 1$, Jensen's inequality gives $\sum_{i \in A} p'(i)^q \le k\big(\frac{1}{N} + \frac{\gamma}{k}\big)^q$ and $\sum_{i \notin A} p'(i)^q \le (N-k)\big(\frac{1}{N} - \frac{\gamma}{N-k}\big)^q$. Therefore, $\sum_i \psi_i p'(i)^q \le \mathrm{PS}_q(N, k, \gamma)$, so $\mathrm{H}_q^{\mathrm{T}}(p') \le \mathrm{H}_q^{\mathrm{T}}\big(p^{(k)}\big)$. Thus, an optimal solution is attained by some distribution of the form $p^{(k)}$ in Equation (34), where $k \le N(1-\gamma)$ ensures entry-wise non-negativity.

$$= \frac{1}{t} \cdot \left(1 + \frac{\xi}{t}\right)^{q-2} + \frac{1}{1-t} \cdot \left(1 - \frac{\xi}{1-t}\right)^{q-2} \tag{39b}$$

$$\geq \frac{1}{t} \cdot \left(1 + \frac{1}{t}\right)^{-2} + \frac{1}{1-t} = \frac{3t+1}{(1-t)(t+1)^2} := G(t). \tag{39c}$$

Here, the inequality in the third line follows from

$$\left(1 + \frac{\xi}{t}\right)^{q-2} \geq \left(1 + \frac{1}{t}\right)^{q-2} \geq \left(1 + \frac{1}{t}\right)^{-2} \quad \text{and} \quad \left(1 - \frac{\xi}{1-t}\right)^{q-2} \geq 1^{q-2} = 1,$$

which are consequences of the monotonicity of $f(x) = x^{q-2}$ for all $0 < q < 1$.

Finally, Equation (38) follows from the bound in Equation (39) because $G(t) \geq 1$, which holds since $(3t+1) - (1-t)(t+1)^2 = t(t^2+t+2) \geq 0$ for all $t \geq 0$, and thus the proof is complete. □

Lemma 5.9 (Upper bound on q-Tsallis entropy via closeness to uniform distribution for $q = 1/2$)**.** *Let p be a probability distribution over $[N]$, where $N \geq 2$, and let ν denote the uniform distribution over $[N]$. Then, for $q = 1/2$, the following inequality holds*:

$$\mathrm{H}_q^{\mathrm{T}}(p) \leq \ln_q\big(N(1 - \mathrm{TV}(p,\nu)^2)\big).$$

Proof. We adopt the same setting as in the proof of Lemma 5.8. To establish the desired upper bound, it therefore suffices to consider the optimization problem in Equation (33), which, by the same argument as in Footnote 14, admits an optimal solution of the form given in Equation (34). Analogously, it remains to bound the objective function, subject to the constraints in Equation (35), which for $q = 1/2$ becomes:

$$\begin{aligned}\mathrm{PS}(N,k,\gamma) := \mathrm{PS}_{1/2}(N,k,\gamma) &= k \cdot \sqrt{\frac{1}{N} + \frac{\gamma}{k}} + (N-k) \cdot \sqrt{\frac{1}{N} - \frac{\gamma}{N-k}} \\ &= \frac{1}{\sqrt{N}}\Big(\underbrace{\sqrt{k(k+\gamma N)}}_{F_0(k;\gamma,N)} + \underbrace{\sqrt{(N-k)((1-\gamma)N-k)}}_{F_1(k;\gamma,N)}\Big).\end{aligned}$$

Define $k_\pm := \frac{N(1-\gamma)}{2} \pm t$, so that $k_+ + k_- = N(1-\gamma)$. It is readily verified that

$$F_0(k_\pm;\gamma,N) = \sqrt{\frac{N(1+\gamma)}{2} \pm t} \cdot \sqrt{\frac{N(1-\gamma)}{2} \pm t} = F_1(k_\mp;\gamma,N).$$

Consequently, $\mathrm{PS}(N,k,\gamma)$ is symmetric about $k = N(1-\gamma)/2$. Moreover, it is straightforward to verify that $\mathrm{PS}(N,k,\gamma)$ is concave in k, since the following holds:

$$\begin{aligned}\frac{\partial^2}{\partial k^2} F_0(k;\gamma,N) &= -\frac{\gamma^2 N^2}{4(k(k+\gamma N))^{3/2}} < 0,\\ \frac{\partial^2}{\partial k^2} F_1(k;\gamma,N) &= -\frac{\gamma^2 N^2}{4((N-k)((1-\gamma)N-k))^{3/2}} < 0.\end{aligned}$$

Therefore, we conclude that the optimal solution of Equation (35) occurs at either $k_* = \lfloor N(1-\gamma)/2 \rfloor$ or $k_* = \lceil N(1-\gamma)/2 \rceil$. As a consequence,

$$\mathrm{H}_q^{\mathrm{T}}(p_{\max}) \leq 2\left(\mathrm{PS}\left(N, \frac{N(1-\gamma)}{2}, \gamma\right) - 1\right) = 2\Big(\sqrt{N(1-\gamma^2)} - 1\Big) = \ln_{1/2}\Big(\sqrt{N(1-\gamma^2)}\Big).$$ □

5.3.2 NIQSZK hardness results

As a warm-up, we first describe the reduction from QSCMM to $\textsc{TsallisQEA}_{1/2}$:

Lemma 5.10 (QSCMM $\leq$ $\textsc{TsallisQEA}_{1/2}$)**.** *Let ρ be an n-qubit quantum state whose purification can be prepared by an m-qubit quantum circuit Q, as described in Definition 5.4. For any such state ρ, define the threshold function*

$$t(n) \coloneqq \left(2^{n/2}-1\right)\left(1-1/n-2^{-n}\right)+2^{n/2}\cdot\frac{\sqrt{2n-1}}{n}-1. \tag{40}$$

Then, the following statements hold:

$$\forall n \geq 4, \qquad \begin{array}{ll} \mathrm{T}(\rho,(I/2)^{\otimes n}) \leq 1/n & \Rightarrow\ \mathrm{S}^{\mathrm{T}}_{1/2}(\rho) \geq t(n)+1/10, \\ \mathrm{T}(\rho,(I/2)^{\otimes n}) \geq 1-1/n & \Rightarrow\ \mathrm{S}^{\mathrm{T}}_{1/2}(\rho) \leq t(n)-1/10. \end{array}$$

Proof. Given a quantum state ρ, let $\{v_i\}_{i\in[2^n]}$ be an orthonormal basis consisting of eigenvectors of ρ, so that it admits the spectral decomposition $\rho=\sum_{i\in[2^n]}\lambda_i|v_i\rangle\langle v_i|$. Let $p\coloneqq(\lambda_1,\cdots,\lambda_{2^n})$ denote the vector of eigenvalues, which forms a probability distribution of dimension 2^n, satisfying $\mathrm{S}^{\mathrm{T}}_{1/2}(\rho)=\mathrm{H}^{\mathrm{T}}_{1/2}(p)$. Let ν denote the uniform distribution over 2^n elements. Since ρ and $(I/2)^{\otimes n}$ commute, they are simultaneously diagonalizable in the same basis. Hence,

$$\mathrm{T}\big(\rho,(I/2)^{\otimes n}\big)=\mathrm{TV}(p,\nu),$$

and the bounds in Lemmas 5.7 and 5.9 naturally extend to the quantum setting.

We now consider the following two cases:

- In the case where $\mathrm{T}(\rho,(I/2)^{\otimes n})\leq 1/n$, the lower bound from Lemma 5.7 implies

$$\mathrm{S}^{\mathrm{T}}_{1/2}(\rho)\geq \ln_{1/2}\left(2^n\right)\cdot\left(1-\mathrm{T}\big(\rho,(I/2)^{\otimes n}\big)-2^{-n}\right)\geq 2\left(2^{n/2}-1\right)\left(1-\frac{1}{n}-2^{-n}\right)\coloneqq \tau_{\mathtt{Y}}(n).$$

- In the case where $\mathrm{T}(\rho,(I/2)^{\otimes n})\geq 1-1/n$, the upper bound from Lemma 5.9 yields

$$\mathrm{S}^{\mathrm{T}}_{1/2}(\rho)\leq \ln_{1/2}\!\left(2^n\left(1-\mathrm{T}\big(\rho,(I/2)^{\otimes n}\big)^2\right)\right)\leq 2^{\frac{n}{2}+1}\frac{\sqrt{2n-1}}{n}-2\coloneqq \tau_{\mathtt{N}}(n).$$

We then define the threshold function $t(n)\coloneqq\frac{1}{2}(\tau_{\mathtt{Y}}(n)+\tau_{\mathtt{N}}(n))$ and the gap function

$$g(n)\coloneqq\frac{1}{2}(\tau_{\mathtt{Y}}(n)-\tau_{\mathtt{N}}(n))=\underbrace{2^{n/2}}_{G_1(n)}\underbrace{\left(1-\frac{1}{n}-2^{-n}-\frac{\sqrt{2n-1}}{n}\right)}_{G_2(n)}+\underbrace{\frac{1}{n}+2^{-n}}_{G_3(n)}.$$

It suffices to show that $g(n)\geq 1/5$ for all integers $n\geq 4$. Since $G_3(n)$ is monotonically decreasing for $n\geq 4$ and satisfies $\lim_{n\to\infty}G_3(n)=0$, it follows that $G_3(n)\geq 0$ for $n\geq 1$, and thus this term does not contribute to the lower bound. Also, it is straightforward to verify that both $G_1(n)$ and $G_2(n)$ are positive and monotonically increasing for $n\geq 4$. Consequently, we complete the proof by noting that

$$g(n)=G_1(n)G_2(n)+G_3(n)\geq G_1(n)G_2(n)\geq G_1(4)G_2(4)=\frac{11}{4}-\sqrt{7}>\frac{1}{10}. \qquad \square$$

Next, we establish the reduction from QSCMM to $\textsc{TsallisQEA}_q$ for all $q\in(0,1)$:

Lemma 5.11 (QSCMM $\leq$ TsallisQEA$_q$ for $0 < q < 1$)**.** *Let ρ be an n-qubit quantum state whose purification can be prepared by an m-qubit quantum circuit Q, as described in Definition 5.4. For any such state ρ, define the threshold function*

$$t(n;q) \coloneqq \frac{2^{n(1-q)}-1}{2(1-q)}\left(2-\frac{1}{n}-2^{-n}\right)-\frac{q}{4}\cdot 2^{n(1-q)}\left(1-\frac{1}{n}\right)^2. \tag{41}$$

Then, the following statements hold:

$$\forall n \geq \left\lceil \frac{5}{q(1-q)} \right\rceil, \quad \begin{array}{ll} \mathrm{T}(\rho,(I/2)^{\otimes n}) \leq 1/n & \Rightarrow \mathrm{S}_q^{\mathrm{T}}(\rho) \geq t(n;q) + 2^{5/q}q/40, \\ \mathrm{T}(\rho,(I/2)^{\otimes n}) \geq 1-1/n & \Rightarrow \mathrm{S}_q^{\mathrm{T}}(\rho) \leq t(n;q) - 2^{5/q}q/40. \end{array}$$

Proof. Following the same reasoning at the beginning of the proof of Lemma 5.10, the bounds in Lemmas 5.7 and 5.8 also extend naturally to the quantum setting.

We now consider the following two cases:

- In the case where $\mathrm{T}(\rho,(I/2)^{\otimes n}) \leq 1/n$, the lower bound from Lemma 5.7 implies

$$\mathrm{S}_q^{\mathrm{T}}(\rho) \geq \ln_q(2^n)\cdot\left(1-\mathrm{T}\big(\rho,(I/2)^{\otimes n}\big)-2^{-n}\right) \geq \frac{2^{n(1-q)}-1}{1-q}\left(1-\frac{1}{n}-2^{-n}\right) \coloneqq \tau_{\mathtt{Y}}(n;q).$$

- In the case where $\mathrm{T}(\rho,(I/2)^{\otimes n}) \geq 1-1/n$, the upper bound from Lemma 5.8 yields

$$\mathrm{S}_q^{\mathrm{T}}(\rho) \leq \ln_q(2^n) - \frac{q}{2}2^{n(1-q)}\cdot\mathrm{T}\big(\rho,(I/2)^{\otimes n}\big)^2 \leq \frac{2^{n(1-q)}-1}{1-q}-\frac{q}{2}\cdot 2^{n(1-q)}\left(1-\frac{1}{n}\right)^2 \coloneqq \tau_{\mathtt{N}}(n;q).$$

We then define the threshold function $t(n;q) \coloneqq \frac{1}{2}(\tau_{\mathtt{Y}}(n;q)+\tau_{\mathtt{N}}(n;q))$ and the gap function

$$\begin{aligned} 2\cdot g(n;q) &\coloneqq \tau_{\mathtt{Y}}(n;q)-\tau_{\mathtt{N}}(n;q) \\ &= \frac{1}{1-q}\cdot\left(\frac{1}{n}+2^{-n}\right)-\frac{2^{n(1-q)}}{1-q}\left(\frac{1}{n}+2^{-n}\right)+\frac{q}{2}\cdot 2^{n(1-q)}\left(1-\frac{1}{n}\right)^2 \\ &\geq \underbrace{\frac{1}{1-q}\cdot\left(\frac{1}{n}+2^{-n}\right)}_{G_1(n;q)}+\underbrace{2^{n(1-q)}}_{G_2(n;q)}\underbrace{\left(\frac{q}{2}\left(1-\frac{1}{n}\right)^2-\frac{2}{n(1-q)}\right)}_{G_3(n;q)}. \end{aligned}$$

Here, the inequality in the third line follows from the facts that $2^{n(1-q)}/(1-q) > 0$ for all $q \in (0,1)$ and $2^{-n} \leq 1/n$ for all $n \geq 1$.

Since $G_1(n;q)$ is non-negative and monotonically decreasing for $n \geq 1$, this term does not contribute to the lower bound. Moreover, as $G_2(n;q) \geq 0$ for all $n \geq 1$ and $q \in (0,1)$, it suffices to show that $G_3(n;q) > 0$ for sufficiently large n.

Since $n^2 G_3(n;q)$ can be written as a quadratic function of n,

$$n^2 G_3(n;q) = \frac{q}{2}\cdot n^2 - \left(q+\frac{2}{1-q}\right)\cdot n + \frac{q}{2},$$

whose sign coincides with that of $G_3(n;q)$, it is evident that $n^2 G_3(n;q) > 0$ whenever

$$n \geq n_0 \coloneqq 1 + \frac{2}{q(1-q)}\left(1+\sqrt{1+q-q^2}\right).$$

Noting that $\sqrt{1+q-q^2} \in (1, \sqrt{5}/2]$ for $q \in (0,1)$, we may safely choose $n \geq n_\star := \frac{5}{q(1-q)} > n_0$ for all $q \in (0,1)$, so that

$$G_3(n;q) \geq G_3(n_\star, q) = \frac{q}{2}\left(1 - \frac{q(1-q)}{5}\right)^2 - \frac{2}{5}q \geq \frac{q}{2}\left(1-\frac{1}{20}\right)^2 - \frac{2}{5}q = \frac{41}{800}\cdot q > \frac{q}{20}.$$

Hence, the gap function satisfies the desired lower bound:

$$g(n;q) \geq \frac{1}{2}G_2(n;q)G_3(n;q) = \frac{1}{2}\cdot 2^{5/q}\cdot\frac{q}{20} = \frac{2^{5/q}q}{40}. \qquad \square$$

As a consequence, Lemma 5.11 implies the following NIQSZK-hardness result:

Theorem 5.5 (TsallisQEA$_q$ is NIQSZK-hard for $0 < q < 1$)**.** *For all real $q \in (0,1)$ and all integers $n \geq \lceil \frac{5}{q(1-q)} \rceil$, the following holds:*

$$\forall g(n;q) \in \left[\frac{1}{\mathrm{poly}(n)}, \frac{2^{5/q}q}{40}\right], \text{ TsallisQEA}_q[t(n;q), g(n;q)] \textit{ is } \textsf{NIQSZK}\textit{-hard.}$$

The threshold parameter $t(n;q)$ is as defined in Equation (41).

Proof. Using Lemma 2.18, we know that QSCMM$[1/n, 1-1/n]$ is NIQSZK-hard for $n \geq 3$. Combining this result with the reduction from QSCMM to TsallisQEA$_q$ for $q \in (0,1)$ and $n \geq \lceil \frac{5}{q(1-q)} \rceil$ (see Lemma 5.11) and the specific choice of $t(n;q)$ in the reduction, we conclude that the resulting gap satisfies $g(n;q) \geq 2^{5/q}q/40$, completing the proof. $\square$

Directly analogous to Theorem 5.5, the reduction in Lemma 5.10 implies the following:

Theorem 5.12 (TsallisQEA$_{1/2}$ is NIQSZK-hard)**.** *For all integers $n \geq 4$, it holds that:*

$$\forall g(n) \in [1/\mathrm{poly}(n), 1/5], \quad \text{TsallisQEA}_{1/2}[t(n), g(n)] \textit{ is } \textsf{NIQSZK}\textit{-hard.}$$

The threshold parameter $t(n)$ is as defined in Equation (40).

References

[ACMT+07] 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. 9

[AISW20] 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. 3, 5, 6, 7, 8

[AK01] 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. 7

[ALL22] 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. 7

[AOST17] 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. 7

[BCWdW01] 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. 3, 4, 5, 7

[Bec01] Christian Beck. Dynamical foundations of nonextensive statistical mechanics. *Physical Review Letters*, 87(18):180601, 2001. doi:10.1103/PhysRevLett.87.180601. 3

[Bec02] 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. 3

[BMW16] Mohammad Bavarian, Saeed Mehraban, and John Wright. Learning entropy. A manuscript on von Neumann entropy estimation, private communication, 2016. 3, 6, 8

[BOW19] Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 503–514, 2019. doi:10.1145/3313276.3316344. 7, 8

[Bru04] Todd A. Brun. Measuring polynomial functions of states. *Quantum Information and Computation*, 4(5):401–408, 2004. doi:10.26421/QIC4.5-6. 7

[BST10] Avraham Ben-Aroya, Oded Schwartz, and Amnon Ta-Shma. Quantum expanders: motivation and construction. *Theory of Computing*, 6(3):47–79, 2010. Preliminary version in *CCC 2008*. doi:10.4086/toc.2010.v006a003. 3, 6, 14

[CCHL21] 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. 8

[CCKV08] André Chailloux, Dragos Florin Ciocan, Iordanis Kerenidis, and Salil Vadhan. Interactive and noninteractive zero knowledge are equivalent in the help model. In *Proceedings of the Fifth Theory of Cryptography Conference*, pages 501–534. Springer, 2008. IACR ePrint:2007/467. doi:10.1007/978-3-540-78524-8_28. 3, 4, 6, 7, 14

[CHLL22] 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. 8

[CHW07] 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. 4, 5, 6, 7, 12

[CW25] Kean Chen and Qisheng Wang. Improved sample upper and lower bounds for trace estimation of quantum state powers. In *Proceedings of the 38th Annual Conference on*

Learning Theory, pages 1008–1028, 2025. URL: https://proceedings.mlr.press/v291/chen25d.html. 1, 3

[CWLY23] 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. 3, 4, 5, 6, 8

[CWYZ25] Kean Chen, Qisheng Wang, Zhan Yu, and Zhicheng Zhang. Simultaneous estimation of nonlinear functionals of a quantum state. ArXiv e-prints, 2025. arXiv:2505.16715. 7

[EAO+02] 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. 3, 4, 5, 7

[EGH+11] 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. 10

[FH13] 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. 10, 11

[GHS21] Tom Gur, Min-Hsiu Hsieh, and Sathyawageeswar Subramanian. Sublinear quantum algorithms for estimating von Neumann entropy. ArXiv e-prints, 2021. arXiv:2111.11139. 7

[GHYZ24] Weiyuan Gong, Jonas Haferkamp, Qi Ye, and Zhihan Zhang. On the sample complexity of purity and inner product estimation. ArXiv e-prints, 2024. arXiv:2410.12712. 3, 4, 5, 6, 8

[GL20] András Gilyén and Tongyang Li. Distributional property testing in a quantum world. In *Proceedings of the 11th Innovations in Theoretical Computer Science Conference*, pages 25:1–25:19, 2020. doi:10.4230/LIPIcs.ITCS.2020.25. 7

[GP22] András Gilyén and Alexander Poremba. Improved quantum algorithms for fidelity estimation. ArXiv e-prints, 2022. arXiv:2203.15993. 3, 8

[GSLW19] 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. 4

[Hay16] Masahito Hayashi. *Quantum Information Theory: Mathematical Foundation*. Cambridge University Press, 2016. doi:10.1007/978-3-662-49725-8. 9

[Hay25] Masahito Hayashi. Measuring quantum relative entropy with finite-size effect. *Quantum*, 9:1725, 2025. doi:10.22331/q-2025-05-05-1725. 5, 8, 11

[HHJ+17] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. doi:10.1109/TIT.2017.2719044. 8, 9

[Hoe63] 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. 13

[Höl89] 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. 9

[JST17] 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. 7

[JVHW15] 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. 6, 7

[JVHW17] 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. 6, 7

[KLN19] Hirotada Kobayashi, François Le Gall, and Harumichi Nishimura. Generalized quantum Arthur–Merlin games. *SIAM Journal on Computing*, 48(3):865–902, 2019. Preliminary version in *CCC 2015*. arXiv:1312.4673, doi:10.1137/17m1160173. 4, 7, 26

[Kob03] Hirotada Kobayashi. Non-interactive quantum perfect and statistical zero-knowledge. In *Proceedings of the 14th International Symposium on Algorithms and Computation*, pages 178–188. Springer, 2003. arXiv:quant-ph/0207158, doi:10.1007/978-3-540-24587-2_20. 3, 14

[KS16] 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. 5

[LA24] Yuhan Liu and Jayadev Acharya. Quantum state testing with restricted measurements. ArXiv e-prints, 2024. arXiv:2408.17439. 8

[Liu26] Yupan Liu. Computational hardness of estimating quantum entropies via binary entropy bounds. In *Proceedings of the 43rd International Symposium on Theoretical Aspects of Computer Science (STACS 2026)*, volume 364 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. arXiv:2601.03734, doi:10.4230/LIPIcs.STACS.2026.66. 4, 6

[LW25] 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. 3, 4, 5, 6, 7, 8, 25, 26, 27

[MdW16] 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. 3, 5, 6, 8, 11

[NS09] 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. 9

[OW16] 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. 8, 12, 13

[OW17] 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. 3, 6, 12

[OW21] 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. 3, 6, 7, 8, 12

[Pan03] Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003. doi:10.1162/089976603321780272. 7

[Pan04] 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. 7

[PTTW26] Angelos Pelecanos, Xinyu Tan, Ewin Tang, and John Wright. Beating full state tomography for unentangled spectrum estimation. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 3313–3363, 2026. doi:10.1137/1.9781611978971.120. 8

[QKW24] 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. 7

[Rén61] 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. 3

[Rog88] 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. 9

[SCC19] 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. 7

[SH21] Sathyawageeswar Subramanian and Min-Hsiu Hsieh. Quantum algorithm for estimating α-renyi entropies of quantum states. *Physical Review A*, 104(2):022428, 2021. doi:10.1103/PhysRevA.104.022428. 7

[SJW$^+$25] Xiao Shi, Jiyu Jiang, Xian Wu, Jingu Xie, Hongshun Yao, and Xin Wang. Near-optimal simultaneous estimation of quantum state moments. ArXiv e-prints, 2025. arXiv:2509.24842. 7

[SLLJ25] Myeongjin Shin, Junseo Lee, Seungwoo Lee, and Kabgyun Jeong. Resource-efficient algorithm for estimating the trace of quantum state powers. *Quantum*, 9:1832, 2025. doi:10.22331/q-2025-08-27-1832. 7

[Tsa88] Constantino Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988. doi:10.1007/BF01016429. 3

[Vaj70] Igor Vajda. Note on discrimination information and variation. *IEEE Transactions on Information Theory*, 16(6):771–773, 1970. doi:10.1109/TIT.1970.1054557. 7

[vEB12] 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. 7

[VV11a] 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, 2011. doi:10.1145/1993636.1993727. 7

[VV11b] 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, 2011. doi:10.1109/FOCS.2011.81. 7

[VV17] 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. 7

[Wan25] Qisheng Wang. Information-theoretic lower bounds for approximating monomials via optimal quantum Tsallis entropy estimation. ArXiv preprint, 2025. arXiv:2509.03496. 7

[Wat02] John Watrous. Limits on the power of quantum statistical zero-knowledge. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 459–468. IEEE, 2002. arXiv:quant-ph/0202111, doi:10.1109/SFCS.2002.1181970. 3

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. 9

[WGL$^+$24] 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, 2024. doi:10.1109/TIT.2024.3399014. 3, 5, 7

[Wil13] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013. doi:10.1017/CBO9781139525343. 9

[WY16] 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. 7

[WZ24a] Qisheng Wang and Zhicheng Zhang. Fast quantum algorithms for trace distance estimation. *IEEE Transactions on Information Theory*, 70(4):2720–2733, 2024. doi:10.1109/TIT.2023.3321121. 5, 8

[WZ24b] Qisheng Wang and Zhicheng Zhang. Sample-optimal quantum estimators for pure-state trace distance and fidelity via samplizer. ArXiv e-prints, 2024. arXiv:2410.21201. 7

[WZ25a] Qisheng Wang and Zhicheng Zhang. Quantum lower bounds by sample-to-query lifting. *SIAM Journal on Computing*, 54(5):1294–1334, 2025. doi:10.1137/24M1638616. 4

[WZ25b] Qisheng Wang and Zhicheng Zhang. Time-efficient quantum entropy estimator via samplizer. *IEEE Transactions on Information Theory*, 71(12):9569–9599, 2025. doi:10.1109/TIT.2025.3576137. 3, 4, 5, 8

[WZL24] 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, 2024. doi:10.1109/TIT.2024.3382037. 3, 5, 7

[YLLW24] Hongshun Yao, Yingjian Liu, Tengxiang Lin, and Xin Wang. Sample-efficient estimation of nonlinear quantum state functions. ArXiv e-prints, 2024. arXiv:2412.01696. 7

[YS21] Justin Yirka and Yiğit Subaşı. Qubit-efficient entanglement spectroscopy using qubit resets. *Quantum*, 5:535, 2021. doi:10.22331/q-2021-09-02-535. 7

[ZL24] 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. 7