Simultaneous Estimation of Nonlinear Functionals of a Quantum State

Kean Chen [*] Qisheng Wang [†] Zhan Yu [‡] Zhicheng Zhang [§]

Abstract

We consider a fundamental task in quantum information theory, estimating the values of $\mathrm{tr}(\mathcal{O}\rho)$, $\mathrm{tr}(\mathcal{O}\rho^2)$, $\ldots$, $\mathrm{tr}(\mathcal{O}\rho^k)$ for an observable $\mathcal{O}$ and a quantum state ρ. We show that $\widetilde{\Theta}(k)$ samples of ρ are sufficient and necessary to simultaneously estimate all the k values. This means that estimating all the k values is almost as easy as estimating only one of them, $\mathrm{tr}(\mathcal{O}\rho^k)$. As an application, our approach advances the sample complexity of entanglement spectroscopy and the virtual cooling for quantum many-body systems. Moreover, we extend our approach to estimating general functionals by polynomial approximation.

[*]Kean Chen is with the Department of Computer and Information Science, University of Pennsylvania, United States (e-mail: keanchen.gan@gmail.com).

[†]Qisheng Wang is with the School of Informatics, University of Edinburgh, Edinburgh, United Kingdom (e-mail: QishengWang1994@gmail.com).

[‡]Zhan Yu is with the Centre for Quantum Technologies, National University of Singapore, Singapore (e-mail: yu.zhan@u.nus.edu).

[§]Zhicheng Zhang is with the Centre for Quantum Software and Information, University of Technology Sydney, Sydney, Australia (e-mail: iszczhang@gmail.com).

1 Introduction

Estimating the expectation value, $\mathrm{tr}(\mathcal{O}\rho)$, of an observable $\mathcal{O}$ with respect to a quantum state ρ is a fundamental task in quantum physics, with a wide range of applications in, e.g., computation with one clean qubit [KL98, She06, CM18], approximating the Jones polynomial [AJL09, SJ08], fidelity estimation [FL11, WZC+23, GP22], quantum state tomography [HHJ+17, OW16, OW17], shadow tomography [Aar18, HKP20, BO21], and quantum machine learning [HKP21]. To estimate expectation values with respect to thermal many-body states at lower temperature, high-order functionals $\mathrm{tr}(\mathcal{O}\rho^k)$ play a key role in the virtual cooling protocol [CCL+19, DTE+25]. These types of high-order functionals were also involved in the virtual distillation protocol [HMO+21, LLWF23] and localized virtual purification protocol [HEY+24].

In this paper, we consider the task of estimating the values of

$$\mathrm{tr}(\mathcal{O}\rho),\ \mathrm{tr}(\mathcal{O}\rho^2),\ \ldots,\ \mathrm{tr}(\mathcal{O}\rho^k) \tag{1}$$

for an observable $\mathcal{O}$ and a quantum state ρ. A direct solution to this task is to estimate each $\mathrm{tr}(\mathcal{O}\rho^j)$ one by one. Specifically, each $\mathrm{tr}(\mathcal{O}\rho^j)$ can be estimated to within additive error ε using $O(j)$ samples of ρ by the generalized SWAP test (by cyclic shift) [EAO+02, Bru04] (see also [CCL+19, HMO+21, SLLJ25]). By individually estimating $\mathrm{tr}(\mathcal{O}\rho^j)$ each with probability $1 - 1/(3k)$ using $O(j\log(k))$ samples of ρ, one can estimate all the k values with probability at least $2/3$ using $\sum_{j=1}^{k} O(j\log(k)) = O(k^2\log(k))$ samples of ρ.

In sharp contrast, we show that all the k values in Equation (1) can be simultaneously estimated with little overhead when it is sufficient to estimate only one of them, $\mathrm{tr}(\mathcal{O}\rho^k)$. We formally state this discovery as follows.

Theorem 1.1 (Simultaneous estimator, Theorems 2.1 and 3.1 combined)**.** *For any known observable $\mathcal{O}$, we can simultaneously estimate $\mathrm{tr}(\mathcal{O}\rho)$, $\mathrm{tr}(\mathcal{O}\rho^2)$, $\ldots$, $\mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error ε using $O(k\log(k)\|\mathcal{O}\|^2/\varepsilon^2)$ samples of ρ, where $\|\cdot\|$ is the operator norm.*

More specifically, given n samples of ρ, we can obtain n random variables $p_1, p_2, \ldots, p_n$ such that for any $1 \le k \le n$,

$$\mathbf{E}[p_k] = \mathrm{tr}(\mathcal{O}\rho^k), \qquad \mathbf{Var}[p_k] \le \frac{2\|\mathcal{O}\|^2 k}{n}.$$

Our estimator is actually optimal up to a logarithmic factor, as shown by the following lower bound.

Theorem 1.2 (Lower bound on estimating a single term, Theorem 4.4 restated)**.** *For any observable $\mathcal{O}$, estimating $\mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error ε requires $\Omega(k\|\mathcal{O}\|^2/\varepsilon^2)$ samples of ρ.*

Roughly speaking, Theorem 1.1 means that we can estimate all the k values in Equation (1) using only $O(k\log(k))$ samples of ρ, quadratically improving the conventional $O(k^2\log(k))$, whereas estimating the single value $\mathrm{tr}(\mathcal{O}\rho^k)$ already requires $\Omega(k)$ samples of ρ due to Theorem 1.2.

As an implication, we can estimate $\mathrm{tr}(\mathcal{O}f(\rho))$ for any polynomial $f \in \mathbb{R}[x]$.

Corollary 1.3 (Special case of Corollary 3.2)**.** *For any known observable $\mathcal{O}$ and polynomial $f \in \mathbb{R}[x]$ of degree k, we can estimate $\mathrm{tr}(\mathcal{O}f(\rho))$ to within additive error ε using $O(k\|\mathcal{O}\|^2\|f\|_1^2/\varepsilon^2)$ samples of ρ, where the ℓ_1-norm $\|f\|_1$ is the absolute sum of the coefficients of f.*

The task of estimating $\mathrm{tr}(\mathcal{O}f(\rho))$ in Corollary 1.3 is an extension of estimating $\mathrm{tr}(\mathcal{O}\rho^k)$ with $f(x) = x^k$. The sample complexity in Corollary 1.3 is optimal up to a constant factor with the hard instance where $f(x) \propto x^k$ by Theorem 1.2. For the special case of $\mathcal{O} = I$, Corollary 1.3 improves

Quantities	Sample Complexity	
	Prior Works	This Work
$\mathrm{tr}(\rho^2)$	$O(1/\varepsilon^2)$ [BCWdW01] $\Omega(1/\varepsilon^2)$ [CWLY23, GHYZ24]	/
$\mathrm{tr}(\rho^k),\ k \geq 3$	$O(k/\varepsilon^2)$ [EAO+02] $\Omega(1/\varepsilon)$ [LW25] $\Omega(1/\varepsilon^2)$ [CW25]	$\Omega(k/\varepsilon^2)$
$\mathrm{tr}(f(\rho))^{\ddagger}$	$O(k^2\|f\|_1^2/\varepsilon^2)$ [QKW24]	$\Theta(k\|f\|_1^2/\varepsilon^2)$
$\{\mathrm{tr}(\mathcal{O}\rho^j)\}_{j=1}^k$	$\widetilde{O}(k^2\|\mathcal{O}\|^2/\varepsilon^2)$ [SLLJ25]	$\widetilde{\Theta}(k\|\mathcal{O}\|^2/\varepsilon^2)$

Table 1: Sample complexity of estimating nonlinear functionals of ρ. ‡ f is a degree-k polynomial.

the result of $O(k^2\|f\|_1^2/\varepsilon^2)$ in [QKW24]. Corollary 1.3 can be further extended to the case with multiple polynomials, i.e., estimating $\mathrm{tr}(\mathcal{O}f_1(\rho))$, $\mathrm{tr}(\mathcal{O}f_2(\rho))$, $\ldots$, $\mathrm{tr}(\mathcal{O}f_m(\rho))$ given m polynomials f_1, f_2, $\ldots$, f_m. See Section 3.2 for more details.

We compare our results with previous work in Table 1.

1.1 The simultaneous estimators

Now we introduce the main idea of constructing the simultaneous estimators in Theorem 1.1.

Suppose ρ is a quantum state in a finite-dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{L}(\mathcal{H})$ denote the set of linear operators on $\mathcal{H}$. Let $\mathfrak{S}_n$ be the symmetric group on $\{1,\ldots,n\}$ and for each permutation $\pi \in \mathfrak{S}_n$, let U_π be the unitary representation of π on $\mathcal{H}^{\otimes n}$, i.e., $U_\pi|\alpha_1\rangle\cdots|\alpha_n\rangle = |\alpha_{\pi^{-1}(1)}\rangle\cdots|\alpha_{\pi^{-1}(n)}\rangle$. Let Φ be the symmetrization map: $\Phi(M) = \frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} U_\pi M U_\pi^\dagger$, where $M \in \mathcal{L}(\mathcal{H}^{\otimes n})$. Then, given an observable $\mathcal{O} \in \mathcal{L}(\mathcal{H})$, for each $1 \leq k \leq n$, we define:

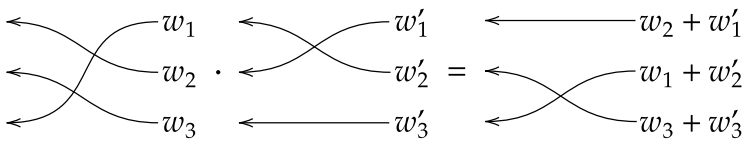
$$\mathcal{O}_k \coloneqq \Phi\big(U_{s_k}\cdot\big(\mathcal{O}\otimes I^{\otimes n-1}\big)\big), \tag{2}$$

where $s_k \in \mathfrak{S}_n$ is the cyclic shift permutation on the first k elements, i.e., $s_k(i) = (i+1) \bmod k$ for $i \leq k$ and $s_k(i) = i$ for $k < i \leq n$. We will show that those $\mathcal{O}_k \in \mathcal{L}(\mathcal{H}^{\otimes n})$ are valid observables and can be used to estimate $\mathrm{tr}(\mathcal{O}\rho^k)$ efficiently.

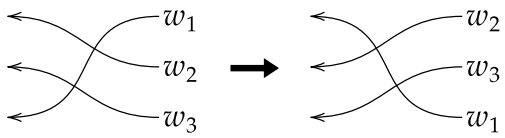
To study the estimators $\mathcal{O}_k$, it is more convenient to treat $U_{s_k}\cdot(\mathcal{O}\otimes I^{\otimes n-1})$ in Equation (2) as a *weighted permutation* (see Definition 2.2). Specifically, a weighted permutation of degree n is a tuple (π, w), where $\pi \in \mathfrak{S}_n$ is a permutation and $w = [w_1,\ldots,w_n] \in \mathbb{Z}_{\geq 0}^n$ is a length-n vector with non-negative integer entries representing the weights. We may directly use πw to denote (π, w) without causing confusion. We define the matrix representation of πw as

$$\mu(\pi w) = U_\pi\cdot(\mathcal{O}^{w_1}\otimes\mathcal{O}^{w_2}\otimes\cdots\otimes\mathcal{O}^{w_n}).$$

Let $\mathfrak{W}_n$ denote the set of all weighted permutations of degree n. Then, $\mathfrak{W}_n \cong \mathfrak{S}_n \ltimes \mathbb{Z}_{\geq 0}^n$ forms a monoid, where the multiplication in $\mathfrak{W}_n$ coincides with the matrix multiplication on its matrix representation, i.e., $\mu(\pi w\cdot\pi' w') = \mu(\pi w)\cdot\mu(\pi' w')$. Additionally, $\mathfrak{W}_n$ is also equipped with an involution "$\dagger$" which coincides with the Hermitian transpose on its matrix representation, i.e., $\mu((\pi w)^\dagger) = \mu(\pi w)^\dagger$. Some examples are shown in Figure 1. We can identify each $\pi \in \mathfrak{S}_n$ as $\pi\mathbf{0} \in \mathfrak{W}_n$ (where $\mathbf{0} \in \mathbb{Z}_{\geq 0}^n$ is the zero vector). Then, we interpret Φ as a symmetrization map acting on the monoid ring $\mathbb{C}\mathfrak{W}_n$ (the set of finite formal sums of elements in $\mathfrak{W}_n$ with complex



(a) Multiplication.



(b) Involution.

Figure 1: Diagrammatic illustration of operations on the weighted permutations.

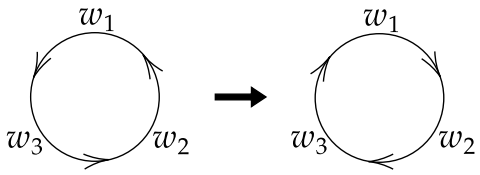


Figure 2: Weighted cycle types before/after the involution.

coefficients), i.e., $\Phi(\mathcal{X}) = \frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} \pi\mathcal{X}\pi^{-1}$ for any $\mathcal{X} \in \mathbb{C}\mathfrak{W}_n$. Thus, our estimators can be written as $\mathcal{O}_k = \mu(\Phi(s_k e_1))$, where $e_1 = [1, 0, \ldots, 0] \in \mathbb{Z}^n_{\geq 0}$.

Then, consider the orbit of a weighted permutation $X \in \mathfrak{W}_n$ under the conjugation action of $\mathfrak{S}_n$, i.e., $\{\pi X\pi^{-1} \,|\, \pi \in \mathfrak{S}_n\}$. It turns out that each orbit corresponds to a generalized cycle type (see Definition 2.10), referred to as a weighted cycle type, where each transition within the cycle is associated with a weight w_i. Figure 2 shows the weighted cycle types corresponding to the example in Figure 1b. We remark that the involution does not necessarily preserve the weighted cycle type in general, which behaves differently from the ordinary cycle types of permutations. Nevertheless, it does preserve the weighted cycle type when the total weight $|w|_1 = \sum_i w_i \leq 2$.

Proposition 1.4 (Lemma 2.11)**.** *For $\pi w \in \mathfrak{W}_n$, if $|w|_1 \leq 2$, then πw and $(\pi w)^\dagger$ have the same weighted cycle type.*

This further leads to the following two key properties of $\mathcal{O}_k$.

1. Hermiticity: The estimator $\mathcal{O}_k$ is Hermitian and thus is a valid observable. Indeed, we have $\mathcal{O}_k^\dagger = \mu(\Phi((s_k e_1)^\dagger))$. Since the total weight $|e_1|_1 = 1 \leq 2$, $(s_k e_1)^\dagger$ and $s_k e_1$ have the same weighted cycle type (thus the same orbit). Note that $\Phi(s_k e_1)$ produces the average of elements in the orbit of $s_k e_1$ and thus equals $\Phi((s_k e_1)^\dagger)$. Therefore, $\mathcal{O}_k^\dagger = \mathcal{O}_k$.

2. Commutativity: The observables $\{\mathcal{O}_i\}_{i=1}^n$ pairwise commute, thereby enabling simultaneous measurements. To prove commutativity, it suffices to show that the coefficient of each weighted permutation in $\mathcal{O}_i\mathcal{O}_j$ coincides with that in $\mathcal{O}_j\mathcal{O}_i = (\mathcal{O}_i\mathcal{O}_j)^\dagger$. Since each weighted permutation occurred in $\mathcal{O}_i\mathcal{O}_j$ has total weight $|w|_1 = 2$, the involution $\dagger$ preserves its weighted cycle type. Therefore, for any weighted cycle type τ, the sum of coefficients of all weighted permutations of type τ in $\mathcal{O}_i\mathcal{O}_j$ is the same as that in $(\mathcal{O}_i\mathcal{O}_j)^\dagger$. On the other hand, since all $\mathcal{O}_i$ are permutation-invariant (i.e., $U_\pi \mathcal{O}_i U_\pi^\dagger = \mathcal{O}_i$ for any $\pi \in \mathfrak{S}_n$), $\mathcal{O}_i\mathcal{O}_j$ is permutation-invariant.

Thus all weighted permutations of type τ must share the same coefficient in $\mathcal{O}_i\mathcal{O}_j$. The same argument also works for $(\mathcal{O}_i\mathcal{O}_j)^\dagger$. Then, we conclude that the coefficient of each weighted permutation in $\mathcal{O}_i\mathcal{O}_j$ is exactly the same as that in $(\mathcal{O}_i\mathcal{O}_j)^\dagger$. More details can be found in Proposition 2.16.

Given the Hermiticity and commutativity of the estimators, our next step is to study the expectation and variance of each estimator. First, it is easy to see that $\mathcal{O}_k$ is an unbiased estimator for $\mathrm{tr}(\mathcal{O}\rho^k)$, i.e., $\mathbf{E}_{\rho^{\otimes n}}[\mathcal{O}_k] = \mathrm{tr}(\mathcal{O}_k\rho^{\otimes n}) = \mathrm{tr}(U_{s_k}(\mathcal{O}\otimes I^{\otimes n-1})\rho^{\otimes n}) = \mathrm{tr}(\mathcal{O}\rho^k)$. Then, to bound the variance, we treat our estimator $\mathcal{O}_k$ as a symmetrization of the natural estimator T_k for the quantity $\mathrm{tr}(\mathcal{O}\rho^k)$, which is

$$T_k = \frac{1}{2\lfloor n/k\rfloor}\sum_{i=0}^{\lfloor n/k\rfloor-1}\Big(\mu(s_{(ik+1,\ldots,ik+k)}e_{ik+1}) + \mu(s_{(ik+1,\ldots,ik+k)}e_{ik+1})^\dagger\Big),$$

where s_J is the cyclic shift permutation on the sequence J and $e_i = [0,\ldots,1,\ldots,0] \in \mathbb{Z}_{\geq 0}^n$ is the vector with a 1 in the i-th place and 0's elsewhere. By the Kadison-Schwarz inequality [Kad52] (cf. [BOW19]), we can show that the variance of the symmetrized estimator $\mathcal{O}_k$ is no more than that of T_k, i.e.,

$$\mathbf{Var}[\mathcal{O}_k] \leq \mathbf{Var}[T_k] \leq \frac{2k\|\mathcal{O}\|^2}{n}.$$

Then, by Chebyshev's inequality and the median trick, it suffices to use $O(k\log(k)\|\mathcal{O}\|^2/\varepsilon^2)$ samples of ρ to give an estimate of $\mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error ε with probability at least $1-1/(3k)$. Similarly, these samples can be reused (due to the commutativity of $\{\mathcal{O}_i\}_{i=1}^k$) to obtain an estimate of each of $\mathrm{tr}(\mathcal{O}\rho^{k-1}),\ldots,\mathrm{tr}(\mathcal{O}\rho)$ with probability $1-1/(3k)$. Therefore, the success probability of the whole process is at least $(1-1/(3k))^k \geq 2/3$.

1.2 Lower bounds

To prove the matching lower bound in Theorem 1.2, we first reduce the problem to the case of $\|\mathcal{O}\| = 1$ and $\langle 0|\mathcal{O}|0\rangle = 1$ by rescaling. In this case, we can obtain a lower bound $\Omega\big(k/\varepsilon^2\big)$ by reducing from the lower bound for quantum state discrimination. Then, we find a new hard instance not well-known in the literature. Specifically, the pair of quantum states ρ_+ and ρ_- for the discrimination task is constructed as:

$$\rho_\pm = \left(1-\frac{1}{k}\pm\frac{\varepsilon}{k}\right)|0\rangle\langle 0| + \left(\frac{1}{k}\mp\frac{\varepsilon}{k}\right)|1\rangle\langle 1|,$$

where $\varepsilon \in (0,1)$. For sufficiently small $\varepsilon > 0$, it can be verified that $\mathrm{tr}(\mathcal{O}\rho_+^k) - \mathrm{tr}(\mathcal{O}\rho_-^k) \geq \Omega(\varepsilon)$. Therefore, any estimator for $\mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error $\Theta(\varepsilon)$ can be used to distinguish ρ_+ and ρ_-. On the other hand, the infidelity between ρ_+ and ρ_- is bounded by $\gamma = 1-\mathrm{F}(\rho_+,\rho_-) \leq O(\varepsilon^2/k)$. By the Helstrom-Holevo bound [Hel67, Hol73] (cf. [Wil13, Hay16]), the sample complexity of distinguishing ρ_+ and ρ_- is lower bounded by $\Omega(1/\gamma) = \Omega(k/\varepsilon^2)$, which easily leads to Theorem 1.2 by rescaling.

Note that another consequence of Theorem 1.2 is the optimality of the generalized SWAP test [EAO+02, Bru04] (see also [CCL+19, HMO+21, SLLJ25]) for estimating a single term $\mathrm{tr}(\mathcal{O}\rho^k)$. In comparison, our Theorem 1.1 implies that simultaneously estimating all the k values in Equation (1) only incurs an $O(\log(k))$ factor.

1.3 Organization of this paper

The construction of the simultaneous estimators in Theorem 1.1 will be presented in Section 2, and their sample complexity will be analyzed in Section 3. The sample complexity lower bounds will be given in Section 4. Applications of our simultaneous estimators will be discussed in Section 5. Finally, a brief discussion will be given in Section 6 with several open questions.

2 Simultaneous Estimators

In this section, we prove the following result.

Theorem 2.1. *Suppose $\mathcal{O} \in \mathcal{L}(\mathcal{H})$ is an observable. Given n samples of an unknown state ρ, there is an algorithm that outputs a list $p_1, p_2, \ldots, p_n$ such that for any $1 \le k \le n$, we have*

$$\mathbf{E}[p_k] = \mathrm{tr}(\mathcal{O}\rho^k) \qquad \text{and} \qquad \mathbf{Var}[p_k] \le \frac{2k\|\mathcal{O}\|^2}{n},$$

where $\|\mathcal{O}\|$ is the operator norm of $\mathcal{O}$.

The proof is based on the idea we outlined in Section 1.1 with full details provided here. First, we formally define the weighted permutations and weighted cycle type in Section 2.1 and Section 2.2. Then, we use these notations to construct our simultaneous estimators in Section 2.3 and bound the variance in Section 2.4. The proof of Theorem 2.1 is summarized in Section 2.5.

2.1 Weighted permutations

First, we define the weighted permutation.

Definition 2.2 (Weighted permutation)**.** *A weighted permutation of degree n is a tuple (π, w) where $\pi \in \mathfrak{S}_n$ is a permutation and $w = [w_1, \ldots, w_n] \in \mathbb{Z}_{\ge 0}^n$ is a length-n vector with non-negative integer entries representing the weights. We use $\mathfrak{W}_n$ to denote the set of all weighted permutations of degree n.*

Then, we define the total weight of a weighted permutation $X \in \mathfrak{W}_n$.

Definition 2.3. *For any $X = (\pi, w) \in \mathfrak{W}_n$, we define the total weight of X as $|X| = |w|_1 = \sum_i w_i$.*

Then, we define the multiplication and involution on $\mathfrak{W}_n$.

Definition 2.4. *We define two operations on $\mathfrak{W}_n$:*

- *Multiplication "$\cdot$":*
$$(\pi, w) \cdot (\pi', w') = (\pi\pi', w_{\pi'} + w'),$$
where $w_{\pi'} = [w_{\pi'(1)}, w_{\pi'(2)}, \ldots, w_{\pi'(n)}] \in \mathbb{Z}_{\ge 0}^n$.
- *Involution "$\dagger$":*
$$(\pi, w)^\dagger = (\pi^{-1}, w_{\pi^{-1}}),$$
where $w_{\pi^{-1}} = [w_{\pi^{-1}(1)}, w_{\pi^{-1}(2)}, \ldots, w_{\pi^{-1}(n)}] \in \mathbb{Z}_{\ge 0}^n$.

Some examples are shown in Figure 1.

Remark 2.1. *For convenience, we may directly use πw to denote the weighted permutation (π, w).*

It is easy to check the following properties.

Fact 2.5. *For any $X, Y, Z \in \mathfrak{W}_n$,*

- $X \cdot (Y \cdot Z) = (X \cdot Y) \cdot Z$,
- $(I\mathbf{0}) \cdot X = X \cdot (I\mathbf{0}) = X$, *where $I \in \mathfrak{S}_n$ is the identity permutation and $\mathbf{0} \in \mathbb{Z}^n_{\geq 0}$ is the zero vector,*
- $(X \cdot Y)^\dagger = Y^\dagger \cdot X^\dagger$.

In fact, $\mathfrak{W}_n$ can be viewed as a monoid $\mathfrak{S}_n \ltimes \mathbb{Z}^n_{\geq 0}$ with an additional involution operation, where $\mathbb{Z}^n_{\geq 0}$ is an abelian monoid with natural addition. On the other hand, we can also see the following properties.

Fact 2.6. *For any $X, Y \in \mathfrak{W}_n$,*

- $|X \cdot Y| = |X| + |Y|$,
- $|X^\dagger| = |X|$.

Then, we consider the matrix representation for $\mathfrak{W}_n$.

Definition 2.7. *Given an Hermitian operator $\mathcal{O} \in \mathcal{L}(\mathcal{H})$, we define the matrix representation $\mu : \mathfrak{W}_n \to \mathcal{L}(\mathcal{H}^{\otimes n})$ based on $\mathcal{O}$ as*

$$\mu(\pi w) = U_\pi \cdot (\mathcal{O}^{w_1} \otimes \mathcal{O}^{w_2} \otimes \cdots \otimes \mathcal{O}^{w_n}),$$

where U_π to denotes the permutation operator acting on the space $\mathcal{H}^{\otimes n}$, i.e., $U_\pi \cdot |\psi_1\rangle \cdots |\psi_n\rangle = |\psi_{\pi^{-1}(1)}\rangle \cdots |\psi_{\pi^{-1}(n)}\rangle$.

We can easily see that μ is a valid matrix representation for $\mathfrak{W}_n$ through the following properties.

Fact 2.8. *For any $X, Y \in \mathfrak{W}_n$,*

- $\mu(X \cdot Y) = \mu(X) \cdot \mu(Y)$,
- $\mu(X)^\dagger = \mu(X^\dagger)$.

Then we extend our definition to the monoid ring $\mathbb{C}\mathfrak{W}_n$, which is the set of formal sums $\sum_{X \in \mathfrak{W}_n} b_X X$, where $b_X \in \mathbb{C}$ and $b_X = 0$ for all but only finitely many X. The involution is extended anti-linearly on $\mathbb{C}\mathfrak{W}_n$, i.e., $(\sum_X b_X X)^\dagger = \sum_X b_X^* X^\dagger$, where b_X^* is the complex conjugate of b_X. The matrix representation μ can be extended naturally on $\mathbb{C}\mathfrak{W}_n$, i.e., $\mu(\sum_X b_X X) = \sum_X b_X \mu(X)$. Then, one can easily check that the properties in Fact 2.5 and Fact 2.8 also hold on $\mathbb{C}\mathfrak{W}_n$. For convenience, we will use the following notation.

Definition 2.9. *For every $X \in \mathfrak{W}_n$, define $c_X : \mathbb{C}\mathfrak{W}_n \to \mathbb{C}$ such that for any $\sum_Y b_Y Y \in \mathbb{C}\mathfrak{W}_n$,*

$$c_X\left(\sum_Y b_Y Y\right) = b_X.$$

2.2 Weighted cycle type

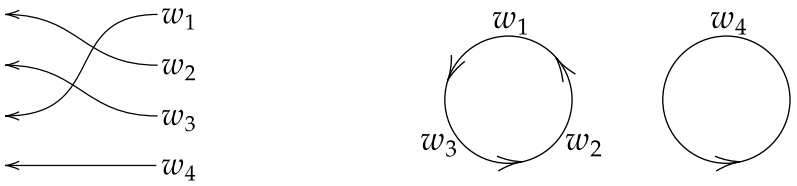
There is a natural inclusion map from $\mathfrak{S}_n$ to $\mathfrak{W}_n$ by identifying each $\pi \in \mathfrak{S}_n$ as $\pi\mathbf{0} \in \mathfrak{M}_n$ (where $\mathbf{0} \in \mathbb{Z}_{\geq 0}^n$ is the zero vector).

Then, consider the conjugation action of $\mathfrak{S}_n$. The orbit of a weighted permutation $X \in \mathfrak{W}_n$ w.r.t. the conjugation action of $\mathfrak{S}_n$ is the set $\{\pi X \pi^{-1} \mid \pi \in \mathfrak{S}_n\}$. Such orbit can be represented by a generalized cycle type, referred to as a weighted cycle type, where each transition within the cycle is associated with a weight w_i.

Definition 2.10 (Weighted cycle type)**.** *A weighted cycle type of degree n is a disjoint union of directed cycle graphs with n vertices, where each edge e is assigned a weight $w_e \in \mathbb{Z}_{\geq 0}$. We use $\tau \vdash n$ to denote that τ is a weighted cycle type of degree n.*

Two weighted cycle types $\tau, \nu \vdash n$ are considered the same if there is a bijection f between the vertex sets of τ and ν such that there exists an edge with weight $w_{v_1 v_2} \in \mathbb{Z}_{\geq 0}$ from v_1 to v_2 if and only if there exists an edge with weight $w_{v_1 v_2}$ from $f(v_1)$ to $f(v_2)$.

An example of weighted cycle type is shown in Figure 3.



(a) A weighted permutation. (b) The weighted cycle type of Figure 3a.

Figure 3: A weighted permutation with the corresponding weighted cycle type.

In general, the involution does not preserve the weighted cycle type of a weighted permutation $X \in \mathfrak{W}_n$ (an example is shown in Figure 2), which behaves differently from the ordinary cycle types of permutations. Nevertheless, it does preserve the weighted cycle type when the total weight $|X| \leq 2$ is small.

Lemma 2.11. *Given a weighted permutation $X \in \mathfrak{W}_n$, if the total weight $|X| \leq 2$, then the weighted cycle type of X is the same as that of $X^\dagger$.*

Proof. Suppose $X = \pi w$. Then, we have $|w|_1 \leq 2$. The involution $\dagger$ only reverses the edge directions in each directed cycle graph. It suffices to prove that a directed cycle graph is invariant under reversing when the sum of its weights is no more than 2.

If the sum of weight is 0, then this is a cycle with all edges having weight 0, and reversing its edge directions gives the same cycle.

If the sum of weight is 1, then this is a cycle with only one edge having weight 1 and others having weight 0, and reversing its edge directions gives the same cycle.

If the sum of weight is 2 with only one edge having weight 2 and others having weight 0, then this is the same as in the previous situation. If there are two edges having weight 1, suppose that, starting from one of these two edges, we need to traverse n_1 weight-0 edges to get to the second weight-1 edge, and then traverse n_2 weight-0 edges to get back to the first weight-1 edge. Then, if the edge directions are reversed, we can start from the second weight-1 edge and traverse n_1 weight-0 edges to get to the first weight-1 edge, and then traverse n_2 weight-0 edges to get back to the second weight-1 edge. Therefore, the reversed cycle graph is the same as the original cycle graph. □

For convenience, we will use the following notation.

Definition 2.12. *For every weighted cycle type $\tau \vdash n$, define $c_\tau \colon \mathbb{C}\mathfrak{W}_n \to \mathbb{C}$ such that for any $\mathcal{X} \in \mathbb{C}\mathfrak{W}_n$,*

$$c_\tau(\mathcal{X}) \coloneqq \sum_{X \text{ has type } \tau} c_X(\mathcal{X}),$$

where $c_X(\cdot)$ is defined in Definition 2.9.

2.3 Simultaneous estimators for $\mathrm{tr}(\mathcal{O}\rho^k)$

Given an observable $\mathcal{O} \in \mathcal{L}(\mathcal{H})$, let μ be the matrix representation of $\mathfrak{W}_n$ based on $\mathcal{O}$ as defined in Definition 2.7. We define the following symmetrization map.

Definition 2.13. *Let Φ be the symmetrization map acting on $\mathbb{C}\mathfrak{W}_n$ as:*

$$\Phi(X) = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \pi X \pi^{-1},$$

where $X \in \mathbb{C}\mathfrak{W}_n$.

Then, our estimators are defined as follows.

Definition 2.14. *For any $1 \le k \le n$, we define*

$$\mathcal{O}_k \coloneqq \mu(\Phi(s_k e_1)), \tag{3}$$

where $s_k \in \mathfrak{S}_n$ is the cyclic shift permutation on the first k elements, i.e., $s_k(i) = (i+1) \bmod k$ for $i \le k$ and $s_k(i) = i$ for $k < i \le n$, and $e_1 = [1, 0, \ldots, 0] \in \mathbb{Z}^n_{\ge 0}$.

We have the following results.

Proposition 2.15. *$\mathcal{O}_k$ is Hermitian and $\mathrm{tr}(\mathcal{O}_k \rho^{\otimes n}) = \mathrm{tr}(\mathcal{O}\rho^k)$.*

Proof. Note that

$$\mathcal{O}_k^\dagger = \mu\big(\Phi(s_k e_1)^\dagger\big) = \mu\big(\Phi\big((s_k e_1)^\dagger\big)\big). \tag{4}$$

Since $|e_1|_1 = 1 \le 2$, by Lemma 2.11, $(s_k e_1)^\dagger$ and $s_k e_1$ have the same weighted cycle type (thus the same orbit). Note that $\Phi(s_k e_1)$ produces the average of elements in the orbit of $s_k e_1$ and thus equals $\Phi((s_k e_1)^\dagger)$. Therefore, $\mathcal{O}_k^\dagger = \mathcal{O}_k$.

On the other hand, we have

$$\begin{aligned}
\mathrm{tr}(\mathcal{O}_k \rho^{\otimes n}) &= \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \mathrm{tr}\left(U_\pi U_{s_k} (\mathcal{O} \otimes I^{\otimes n-1}) U_\pi^\dagger \rho^{\otimes n}\right) \\
&= \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \mathrm{tr}\left(U_{s_k} (\mathcal{O} \otimes I^{\otimes n-1}) \rho^{\otimes n}\right) \\
&= \mathrm{tr}(\mathcal{O}\rho^k).
\end{aligned}$$

□

Proposition 2.16. *For any $1 \le i \le j \le n$,*

$$\mathcal{O}_i \mathcal{O}_j = \mathcal{O}_j \mathcal{O}_i.$$

Proof. By Proposition 2.15, we know that $\mathcal{O}_i^\dagger = \mathcal{O}_i$, which means $\mathcal{O}_j\mathcal{O}_i = (\mathcal{O}_i\mathcal{O}_j)^\dagger$. Let P_i denote $\Phi(s_i e_1)$. Thus, $\mathcal{O}_i\mathcal{O}_j = \mu(P_i)\mu(P_j) = \mu(P_iP_j)$. Then, it suffices to show that

$$P_iP_j = (P_iP_j)^\dagger.$$

Suppose

$$P_iP_j = \sum b_X X,$$

where $X \in \mathfrak{W}_n$ and $b_X = 0$ for all but only finitely many X. From the definition of P_i, it is obvious that b_X are real numbers. Then, $(P_iP_j)^\dagger = \sum b_X X^\dagger$. Furthermore, by the definition of P_i and Fact 2.6, we know that for all X such that $b_X \neq 0$, we have $|X| = 2$, and thus $X^\dagger$ has the same weighted cycle type as X by Lemma 2.11. Therefore, for any weighted cycle type $\tau \vdash n$, we have

$$c_\tau(P_iP_j) = c_\tau((P_iP_j)^\dagger), \tag{5}$$

where $c_\tau(\cdot)$ is defined in Definition 2.12.

On the other hand, since all P_i are permutation-invariant by definition (i.e., $\pi P_i \pi^{-1} = P_i$ for any $\pi \in \mathfrak{S}_n$), P_iP_j is also permutation-invariant. Note that the conjugation action of $\mathfrak{S}_n$ acts transitively on each orbit in $\mathfrak{W}_n$. That is, for any $X, Y \in \mathfrak{W}_n$, if X, Y are in the same orbit (of the same weighted cycle type), there exists a $\pi \in \mathfrak{S}_n$ such that

$$\pi X \pi^{-1} = Y.$$

Further note that the conjugation map $\pi(\cdot)\pi^{-1}$ is injective. We can conclude that X and Y must have the same coefficient in P_iP_j, i.e., $c_X(P_iP_j) = c_Y(P_iP_j)$. Therefore, if X is of type τ, then $c_X(P_iP_j) = c_\tau(P_iP_j)/|\tau|$, where $|\tau|$ is the number of weighted permutations in the orbit τ. Since $(P_iP_j)^\dagger$ is also permutation-invariant, the same argument works for $(P_iP_j)^\dagger$, thus we have $c_X((P_iP_j)^\dagger) = c_\tau((P_iP_j)^\dagger)/|\tau|$. Then, for any $X \in \mathfrak{W}_n$, letting τ be its weighted cycle type, we have

$$c_X((P_iP_j)^\dagger) = c_\tau((P_iP_j)^\dagger)/|\tau| = c_\tau(P_iP_j)/|\tau| = c_X(P_iP_j),$$

where the second equality is by Equation (5). Therefore, $P_iP_j = (P_iP_j)^\dagger$. □

2.4 Bounding the variance

Now, we fix ρ and introduce the following notations. Suppose $\mathcal{Q}$ is an observable acting on $\mathcal{H}^{\otimes n}$, let x be the outcome of measuring $\rho^{\otimes n}$ with the observable $\mathcal{Q}$. Then we write $\mathbf{E}[\mathcal{Q}] \coloneqq \mathbf{E}[x]$ and $\mathbf{Var}[\mathcal{Q}] \coloneqq \mathbf{Var}[x]$. It is easy to see that

$$\mathbf{E}[\mathcal{Q}] = \mathrm{tr}(\mathcal{Q}\rho^{\otimes n}), \quad \mathbf{Var}[\mathcal{Q}] = \mathbf{E}[\mathcal{Q}^2] - \mathbf{E}[\mathcal{Q}]^2 = \mathrm{tr}(\mathcal{Q}^2\rho^{\otimes n}) - \mathrm{tr}(\mathcal{Q}\rho^{\otimes n})^2.$$

Let $\mathcal{O}_k$ be the observables defined in Definition 2.14. We will bound $\mathbf{Var}[\mathcal{O}_k]$ in this section. For a sequence $J = (j_1, j_2, \ldots, j_l) \subseteq [n]$, let $s_J \in \mathfrak{S}_n$ be the cyclic shift permutation acting on J, i.e., s_J maps j_i to j_{i+1}, maps j_l to j_1 and leaves other elements unchanged. Let

$$T_k \coloneqq \frac{1}{2\lfloor n/k \rfloor} \sum_{i=0}^{\lfloor n/k \rfloor - 1} \left(\mu(s_{(ik+1,\ldots,ik+k)} e_{ik+1}) + \mu(s_{(ik+1,\ldots,ik+k)} e_{ik+1})^\dagger \right),$$

where $e_i = [0, \ldots, 1, \ldots, 0] \in \mathbb{Z}_{\geq 0}^n$ is the vector with a 1 in the i-th place and 0's elsewhere. We have the following result

Lemma 2.17.

$$\mathbf{Var}[\mathcal{O}_k] \leq \mathbf{Var}[T_k].$$

Proof. The proof follows the idea in [BOW19, Lemma 3.14].

By symmetry, for any i, we have

$$\Phi(s_{(ik+1,\ldots,ik+k)}e_{ik+1}) = \Phi(s_k e_1),$$

where Φ is the symmetrization map defined in Definition 2.13. Therefore,

$$\Phi(T_k) = \frac{1}{2}(\mathcal{O}_k + \mathcal{O}_k^\dagger) = \mathcal{O}_k,$$

where with a slight abuse of notations, Φ is reinterpreted as the symmetrization map acting on $\mathcal{L}(\mathcal{H})$, i.e., $\Phi(A) = \frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} U_\pi A U_\pi^\dagger$ for any $A \in \mathcal{L}(\mathcal{H}^{\otimes n})$. Note that Φ is both positive and unital, by the Kadison-Schwarz inequality [Kad52], we have

$$\Phi(T_k)^2 \sqsubseteq \Phi(T_k^2),$$

where $\sqsubseteq$ is the Loewner order. This means

$$\begin{aligned}
\mathbf{E}[\mathcal{O}_k^2] &= \mathrm{tr}(\mathcal{O}_k^2\rho^{\otimes n}) = \mathrm{tr}(\Phi(T_k)^2\rho^{\otimes n}) \\
&\leq \mathrm{tr}(\Phi(T_k^2)\rho^{\otimes n}) \\
&= \mathrm{tr}(T_k^2\rho^{\otimes n}) \qquad (6) \\
&= \mathbf{E}[T_k^2],
\end{aligned}$$

where Equation (6) is because $\pi \in \mathfrak{S}_n$ commutes with $\rho^{\otimes n}$. Since $\mathbf{E}[\mathcal{O}_k] = \mathrm{tr}(\rho^k) = \mathbf{E}[T_k]$, we have $\mathbf{Var}[\mathcal{O}_k] \leq \mathbf{Var}[T_k]$. □

Then, it remains to upper bound $\mathbf{Var}[T_k]$. We have the following result.

Lemma 2.18.

$$\mathbf{Var}[T_k] \leq \frac{2k\|\mathcal{O}\|^2}{n}.$$

Proof. Let S_i denote $\mu(s_{(ik+1,\ldots,ik+k)}e_{ik+1}) + \mu(s_{(ik+1,\ldots,ik+k)}e_{ik+1})^\dagger$. Note that for $i \neq j$, S_i and S_j act non-trivially on different subsystems, and thus

$$\mathbf{E}[S_i S_j] = \mathbf{E}[S_i]\,\mathbf{E}[S_j].$$

Therefore

$$\begin{aligned}
\mathbf{Var}[T_k] &= \mathbf{Var}\left[\frac{1}{2\lfloor n/k\rfloor}\sum_{i=0}^{\lfloor n/k\rfloor-1} S_i\right] \\
&= \frac{1}{4\lfloor n/k\rfloor^2}\sum_{i=0}^{\lfloor n/k\rfloor-1}\mathbf{Var}[S_i] \\
&\leq \frac{1}{4\lfloor n/k\rfloor^2}\cdot\lfloor n/k\rfloor\cdot 4\|\mathcal{O}\|^2 \qquad (7) \\
&= \frac{\|\mathcal{O}\|^2}{\lfloor n/k\rfloor} \\
&\leq \frac{2k\|\mathcal{O}\|^2}{n},
\end{aligned}$$

where in Equation (7) we use the fact that the operator norm of S_i is upper bounded by $2\|\mathcal{O}\|$. □

Then, we have the following result.

Proposition 2.19.

$$\mathbf{Var}[\mathcal{O}_k] \le \frac{2k\|\mathcal{O}\|^2}{n}.$$

Proof. This is by combining Lemma 2.17 with Lemma 2.18. □

2.5 Proof of Theorem 2.1

Proof of Theorem 2.1. We define $p_1, \ldots, p_n$ to be the outcomes of measuring $\rho^{\otimes n}$ with the observables $\mathcal{O}_1, \ldots, \mathcal{O}_n$. Note that this is well-defined as $\mathcal{O}_i$ are valid observables and pairwise commuting due to Proposition 2.15 and Proposition 2.16. Thus, we can perform the measurements simultaneously. Then, by Proposition 2.15 and Proposition 2.19, we have $\mathbf{E}[p_k] = \mathbf{E}[\mathcal{O}_k] = \mathrm{tr}(\mathcal{O}\rho^k)$ and $\mathbf{Var}[p_k] = \mathbf{Var}[\mathcal{O}_k] \le 2k\|\mathcal{O}\|^2/n$. □

3 Sample Complexity Upper Bounds

3.1 Nonlinear functionals

Theorem 3.1. *For any known observable $\mathcal{O}$, we can simultaneously estimate $\mathrm{tr}(\mathcal{O}\rho), \mathrm{tr}(\mathcal{O}\rho^2), \ldots, \mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error ε using $O(k\log(k)\|\mathcal{O}\|^2/\varepsilon^2)$ samples of ρ, where $\|\cdot\|$ is the operator norm.*

Proof. From Theorem 2.1, given n samples of an unknown state ρ, we can outputs a list $p_1, p_2, \ldots, p_n$ such that for any $1 \le k \le n$,

$$\mathbf{E}[p_k] = \mathrm{tr}(\mathcal{O}\rho^k) \qquad \text{and} \qquad \mathbf{Var}[p_k] \le \frac{2k\|\mathcal{O}\|^2}{n}.$$

Using Chebyshev's inequality, we have

$$\Pr[|p_k - \mathbf{E}[p_k]| \ge \varepsilon] \le \frac{2k\|\mathcal{O}\|^2}{n\varepsilon^2}.$$

We take $n = \lceil 6k\|\mathcal{O}\|^2/\varepsilon^2 \rceil$ to ensure that the probability

$$\Pr[|p_j - \mathbf{E}[p_j]| \ge \varepsilon] \le \frac{2j\|\mathcal{O}\|^2}{n\varepsilon^2} \le \frac{1}{3}$$

for all $1 \le j \le k$. Then we can use the median trick to further boost the confidence of estimators p_j to at least $1 - 1/(3k)$ for all $1 \le j \le k$ with an $O(\log(k))$ overhead.

Specifically, we repeat the above estimation process for m times and obtain the random variables $p_j^{(1)}, p_j^{(2)}, ..., p_j^{(m)}$ for all $1 \le j \le k$. Then we pick the median $\tilde{p}_j = \mathrm{median}(p_j^{(1)}, p_j^{(2)}, ..., p_j^{(m)})$ as the estimator of $\mathrm{tr}(\mathcal{O}\rho^j)$. Note that each $p_j^{(\ell)}$ for $1 \le \ell \le m$ is an independent random variable such that

$$\Pr\left[\left|p_j^{(\ell)} - \mathrm{tr}(\mathcal{O}\rho^j)\right| \ge \varepsilon\right] \le \frac{1}{3}.$$

Define indicator variables $Y_{j,\ell} = \mathbb{1}_{\{|p_j^{(\ell)} - \mathrm{tr}(\mathcal{O}\rho^j)| \ge \varepsilon\}}$, so $Y_{j,\ell} \sim \mathrm{Bernoulli}(p)$ with $p \le 1/3$. Let $Y_j = \sum_{\ell=1}^m Y_{j,\ell}$ be the number of "bad" estimates among the m trials. To bound the probability that more than half of the $Y_{j,\ell}$'s are bad,

$$\Pr[|\tilde{p}_j - \mathrm{tr}(\mathcal{O}\rho^j)| > \varepsilon] \le \Pr[Y_j > m/2].$$

Now apply Hoeffding's inequality [Hoe63, Theorem 2] with $\mathbf{E}[Y_j] = mp \le m/3$, we have

$$\Pr[Y_j > m/2] = \Pr[Y_j - \mathbf{E}[Y_j] > m/6] \le \exp(-m/18).$$

By choosing $m = \lceil 18 \ln(3k) \rceil$, we have the success probability

$$\Pr\left[\left|\tilde{p}_j - \operatorname{tr}(\mathcal{O}\rho^j)\right| \le \varepsilon\right] \ge 1 - \frac{1}{3k}.$$

Then by the union bound, we can simultaneously estimate $\operatorname{tr}(\mathcal{O}\rho), \operatorname{tr}(\mathcal{O}\rho^2), \ldots, \operatorname{tr}(\mathcal{O}\rho^k)$ to within additive error ε with probability at least $2/3$. The entire estimation process uses $n \cdot m = \lceil 6k\|\mathcal{O}\|^2/\varepsilon^2 \rceil \cdot \lceil 18 \ln(3k) \rceil = O(k \log(k)\|\mathcal{O}\|^2/\varepsilon^2)$ samples of ρ. □

3.2 Extension to general functionals by polynomial approximation

Our method enables us to simultaneously estimate several functionals of the form $\operatorname{tr}(\mathcal{O}g(\rho))$, where g is a real function.

Corollary 3.2. *Let $\mathcal{O}$ be an observable and $g_1, g_2, \ldots, g_m \colon [0,1] \to \mathbb{R}$ that can be approximated respectively by degree-k polynomials $f_1, f_2, \ldots, f_m$ to precision $\varepsilon/2\|\mathcal{O}\|d$. For any d-dimensional state ρ, we can simultaneously estimate $\operatorname{tr}(\mathcal{O}g_1(\rho))$, $\operatorname{tr}(\mathcal{O}g_2(\rho))$, $\ldots$, $\operatorname{tr}(\mathcal{O}g_m(\rho))$ to within additive error ε using*

$$O\left(\frac{k\|\mathcal{O}\|^2 \log(\min\{k,m\})}{\varepsilon^2} \max_{1 \le i \le m} \|f_i\|_1^2\right)$$

samples of ρ, where $\|f_i\|_1$ is the ℓ_1-norm of the polynomial coefficients of f_i.

Corollary 3.2 improves and generalizes the results in [QKW24] (see also [YLLW25]), where they presented an approach for the case where $\mathcal{O} = I$ and $m = 1$ with sample complexity $O(k^2\|f_1\|_1^2/\varepsilon^2)$. In particular, Corollary 3.2 improves it to $O(k\|f_1\|_1^2/\varepsilon^2)$ by a factor of k, and if there are m different functionals to estimate, it incurs only an overhead of $\log(\min\{k,m\})$. Moreover, Corollary 3.2 is optimal even when $m = 1$, $g_1(x) = f_1(x) \propto x^k$, and $\mathcal{O} \propto I$, where $\Omega(k\|\mathcal{O}\|^2\|f_1\|_1^2/\varepsilon^2)$ samples of ρ are required as implied by Theorem 1.2.

Proof of Corollary 3.2. Let us first focus on a single polynomial $f = \sum_{j=1}^k \alpha_j x^j$. For a quantum state ρ, using Theorem 1.1, we can obtain $p_1, \ldots, p_k$ such that for $j = 1$ to k,

$$\mathbf{E}[p_j] = \operatorname{tr}\left(\mathcal{O}\rho^j\right), \text{ and } \mathbf{Var}[p_j] = \frac{2\|\mathcal{O}\|^2 j}{n}.$$

Suppose that $X_1, \ldots, X_k$ are random variables. Observe that the standard variation of the random variable $X = X_1 + \ldots + X_k$ can be upper bounded by the sum of the standard variations of all

$X_1, \ldots, X_k$ as follows:

$$\begin{aligned}
\sigma[X_1 + \ldots + X_k] &= \sqrt{\mathbf{Var}[X_1 + \ldots + X_k]} \\
&= \sqrt{\sum_j \mathbf{Var}[X_j] + \sum_{j \neq l} \mathbf{Cov}[X_j, X_l]} \\
&\leq \sqrt{\sum_j \mathbf{Var}[X_j] + \sum_{j \neq l} \sqrt{\mathbf{Var}[X_j]\,\mathbf{Var}[X_l]}} \\
&= \sqrt{\left(\sum_j \sigma[X_j]\right)^2} \\
&= \sum_j \sigma[X_j],
\end{aligned}$$

where $\mathbf{Cov}[\cdot, \cdot]$ denotes the covariance between two random variables, and we use the Cauchy–Schwarz inequality. Let $\hat{p} = \sum_{j=1}^k a_j p_j$. Then, we can calculate $\mathbf{E}[\hat{p}] = \mathrm{tr}(\mathcal{O} f(\rho))$ and

$$\sigma[\hat{p}] \leq \sum_{j=1}^k a_j \sigma[p_j] = \sum_{j=1}^k \sqrt{\frac{2\|\mathcal{O}\|^2 j}{n}} \alpha_j,$$

by the above observation. Consequently, setting the number of samples

$$n = O\left(\frac{k\|f\|_1^2 \|\mathcal{O}\|^2}{\varepsilon^2}\right)$$

allows us to estimate $\mathrm{tr}(\mathcal{O} f(\rho))$ to within additive error $\varepsilon/2$ and with probability $\geq 2/3$. Like in the proof of Theorem 3.1, by repeating this experiment for $O(\log(1/\delta))$ times, we can amplify the success probability to $1 - \delta$.

Now consider multiple polynomials $f_1, f_2, \ldots, f_m$.

- If $m \leq k$, then for each $i = 1$ to m, we can estimate $\mathrm{tr}(\mathcal{O} f_i(\rho))$ to within additive error $\varepsilon/2$ and with probability $\geq 1 - 1/(3m)$. By the union bound, with overall success probability $\geq 2/3$, we can simultaneously estimate all $\mathrm{tr}(\mathcal{O} f_1(\rho)), \ldots, \mathrm{tr}(\mathcal{O} f_m(\rho))$ to within error $\varepsilon/2$.
- If $m > k$, then for each $j = 1$ to k, we can estimate $\mathrm{tr}\big(\mathcal{O}\rho^j\big)$ to within additive error $\varepsilon/(2\max_{i=1}^m \|f_i\|_1)$ and with probability $\geq 1 - 1/(3k)$. By the union bound, with success probability $\geq 2/3$, for any $i = 1$ to m, the linear combination of estimates of $\mathrm{tr}\big(\mathcal{O}\rho^j\big)$ gives an estimation of $\mathrm{tr}(\mathcal{O} f_i(\rho))$ to within error $\varepsilon/2$.

Combining the results above, we can simultaneously estimate all $\mathrm{tr}(\mathcal{O} f_1(\rho)), \ldots, \mathrm{tr}(\mathcal{O} f_m(\rho))$ to within additive error $\varepsilon/2$, with overall success probability $\geq 2/3$, using

$$O\left(\frac{k\|\mathcal{O}\|^2 \max_{i=1}^m \|f_i\|_1^2 \log(\min\{k, m\})}{\varepsilon^2}\right)$$

samples of ρ. Since in Corollary 3.2, we assume that each $f_i(x)$ approximates $g_i(x)$ to precision $\varepsilon/(2\|\mathcal{O}\|d)$ for $x \in [0, 1]$, with probability $\geq 2/3$, our estimates of $\mathrm{tr}(\mathcal{O} f_i(\rho))$ approximate $\mathrm{tr}(\mathcal{O} g_i(\rho))$ to within error ε. The conclusion immediately follows. □

4 Sample Complexity Lower Bounds

In this section, we first establish a matching lower bound on the sample complexity of estimating $\mathrm{tr}(\rho^k)$. Then, we extend the proof idea to show a lower bound on the sample complexity of estimating $\mathrm{tr}(\mathcal{O}\rho^k)$, where $\mathcal{O}$ is a given observable. Further, we derive lower bounds on the sample complexity of estimating $\mathrm{tr}(f(\rho))$ and $\mathrm{tr}(\mathcal{O}f(\rho))$ for general functional f approximated by polynomials.

4.1 Estimation of trace powers $\mathrm{tr}(\rho^k)$

Let us start with the lower bound on the sample complexity of estimating the trace power $\mathrm{tr}(\rho^k)$; that is, the special case $\mathcal{O} = I$ in Theorem 1.2. Before proceeding, we introduce the following theorem and fact about quantum state discrimination.

Theorem 4.1 (Quantum state discrimination, cf. [Wil13, Section 9.1.4] and [Hay16, Lemma 3.2])**.** *Let ρ_0 and ρ_1 be two quantum states, and let ρ be a quantum state such that $\rho = \rho_0$ or $\rho = \rho_1$ with equal probability. Then, any POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$ determines whether $\rho = \rho_0$ and $\rho = \rho_1$ with success probability at most*

$$\frac{1}{2} + \frac{1}{4}\|\rho_0 - \rho_1\|_1,$$

where

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

is the trace distance.

Fact 4.2. *Any quantum algorithm that distinguishes between two quantum states ρ_0 and ρ_1 requires sample complexity $\Omega(1/\gamma)$, where $\gamma = 1 - \mathrm{F}(\rho_0, \rho_1)$ is the infidelity and*

$$\mathrm{F}(\rho_0, \rho_1) = \mathrm{tr}\left(\sqrt{\sqrt{\rho_0}\rho_1\sqrt{\rho_0}}\right)$$

is the fidelity.

Now we can prove Theorem 1.2 for the special case of $\mathcal{O} = I$.

Theorem 4.3. *For sufficiently large integer k and sufficiently small $\varepsilon > 0$, estimating $\mathrm{tr}(\rho^k)$ to within additive error ε requires sample complexity $\Omega(k/\varepsilon^2)$.*

The lower bound $\Omega(k/\varepsilon^2)$ in Theorem 4.3 considers the dependence on k, whereas the previous lower bound $\Omega(1/\varepsilon^2)$ in [CW25] assumes $k = \Theta(1)$.

Proof of Theorem 4.3. Suppose that estimating $\mathrm{tr}(\rho^k)$ to additive error ε can be done with sample complexity $S(k, \varepsilon)$. Let $\varepsilon \in (0, 1)$. Consider the problem of distinguishing the two quantum states $\rho_\pm$ defined by

$$\rho_\pm = \left(1 - \frac{1}{k} \pm \frac{\varepsilon}{k}\right)|0\rangle\langle 0| + \left(\frac{1}{k} \mp \frac{\varepsilon}{k}\right)|1\rangle\langle 1|. \tag{8}$$

Note that

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

Then, we have

$$\mathrm{tr}(\rho_+^k) - \mathrm{tr}(\rho_-^k) = \Omega(\varepsilon) \tag{9}$$

for sufficiently small $\varepsilon > 0$, which is by noting that

$$\begin{aligned}
\lim_{\varepsilon\to 0}\frac{\mathrm{tr}(\rho_+^k)-\mathrm{tr}(\rho_-^k)}{\varepsilon} &= \frac{\partial}{\partial\varepsilon}\Big(\mathrm{tr}(\rho_+^k)-\mathrm{tr}(\rho_-^k)\Big)\Big|_{\varepsilon=0} \\
&= \left(\left(1-\frac{1}{k}+\frac{\varepsilon}{k}\right)^{k-1}-\left(\frac{1}{k}-\frac{\varepsilon}{k}\right)^{k-1}+\left(1-\frac{1}{k}-\frac{\varepsilon}{k}\right)^{k-1}-\left(\frac{1}{k}+\frac{\varepsilon}{k}\right)^{k-1}\right)\Big|_{\varepsilon=0} \\
&= 2\left(\left(1-\frac{1}{k}\right)^{k-1}-\left(\frac{1}{k}\right)^{k-1}\right) \geq \Theta(1),
\end{aligned}$$

where the last inequality is because

$$\lim_{k\to\infty}\left(\left(1-\frac{1}{k}\right)^{k-1}-\left(\frac{1}{k}\right)^{k-1}\right)=\frac{1}{e}.$$

Then, we can distinguish ρ_+ and ρ_- by separately estimating $\mathrm{tr}(\rho_+^k)$ and $\mathrm{tr}(\rho_-^k)$ to additive error $\Theta(\varepsilon)$, each with sample complexity $S(k,\Theta(\varepsilon))$.

On the other hand, by Fact 4.2, the quantum sample complexity of distinguishing ρ_+ and ρ_- is

$$\Omega\left(\frac{1}{1-\mathrm{F}(\rho_+,\rho_-)}\right)\geq\Omega\left(\frac{k}{\varepsilon^2}\right)$$

for sufficiently small $\varepsilon > 0$, where the inequality is by noting that

$$\begin{aligned}
\lim_{\varepsilon\to 0}\frac{1-\mathrm{F}(\rho_+,\rho_-)}{\varepsilon^2} &= \lim_{\varepsilon\to 0}\frac{1-\sqrt{\left(1-\frac{1}{k}\right)^2-\frac{\varepsilon^2}{k^2}}-\sqrt{\left(\frac{1}{k}\right)^2-\frac{\varepsilon^2}{k^2}}}{\varepsilon^2} \\
&= \lim_{x\to 0}\frac{1-\sqrt{\left(1-\frac{1}{k}\right)^2-\frac{x}{k^2}}-\sqrt{\left(\frac{1}{k}\right)^2-\frac{x}{k^2}}}{x} \\
&= \frac{\partial}{\partial x}\left(1-\sqrt{\left(1-\frac{1}{k}\right)^2-\frac{x}{k^2}}-\sqrt{\left(\frac{1}{k}\right)^2-\frac{x}{k^2}}\right)\Big|_{x=0} \\
&= \left(\frac{1}{2k^2\sqrt{\left(1-\frac{1}{k}\right)^2-\frac{x}{k^2}}}+\frac{1}{2k^2\sqrt{\left(\frac{1}{k}\right)^2-\frac{x}{k^2}}}\right)\Big|_{x=0} \\
&= \frac{1}{2(k-1)}=\Theta\left(\frac{1}{k}\right).
\end{aligned}$$

The above together gives

$$S(k,\Theta(\varepsilon))\geq\Omega\left(\frac{k}{\varepsilon^2}\right)$$

for sufficiently small $\varepsilon > 0$ (dependent on k). By letting $\delta = \Theta(\varepsilon)$, we have

$$S(k,\delta)\geq\Omega\left(\frac{k}{\delta^2}\right).$$

These yield the proof. □

4.2 Estimation of $\mathrm{tr}(\mathcal{O}\rho^k)$ with general observables $\mathcal{O}$

Next, we formally restate Theorem 1.2 in the following theorem and then show a lower bound on the sample complexity of estimating $\mathrm{tr}(\mathcal{O}\rho^k)$ for any given observable $\mathcal{O}$.

Theorem 4.4. *For sufficiently large integer k, sufficiently small $\varepsilon > 0$, and any observable $\mathcal{O}$, estimating $\mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error ε requires sample complexity $\Omega(k\|\mathcal{O}\|^2/\varepsilon^2)$.*

Proof. Estimating $\mathrm{tr}(\mathcal{O}\rho^k)$ to within additive error ε is equivalent to estimating $\mathrm{tr}(\frac{\mathcal{O}}{\|\mathcal{O}\|}\rho^k)$ to within error $\frac{\varepsilon}{\|\mathcal{O}\|}$. So, it suffices to assume $\|\mathcal{O}\| = 1$ and prove that estimating $\mathrm{tr}(\mathcal{O}\rho^k)$ to error ε requires $\Omega(k/\varepsilon^2)$ samples of ρ.

The proof idea is similar to that for Theorem 4.3. Suppose that ρ is a d-dimensional quantum state. Without loss of generality, we set the computational basis $|x\rangle_{x=0}^{d-1}$ such that $\mathcal{O}$ is diagonal in this basis and that $\langle 0|\mathcal{O}|0\rangle = 1$. Suppose that $\langle 1|\mathcal{O}|1\rangle = a$ for some $a \in [-1, 1]$. Like the proof of Theorem 4.3, let us consider the problem of distinguishing the two quantum states $\rho_\pm$ defined in Equation (8). We can calculate

$$\mathrm{tr}\Big(\mathcal{O}\rho_\pm^k\Big) = \left(1 - \frac{1}{k} \pm \frac{\varepsilon}{k}\right)^k + a\left(\frac{1}{k} \mp \frac{\varepsilon}{k}\right)^k.$$

It can be verified that

$$\begin{aligned}
\left|\mathrm{tr}\Big(\mathcal{O}\rho_+^k\Big) - \mathrm{tr}\Big(\mathcal{O}\rho_-^k\Big)\right| &= \left(1 - \frac{1}{k} + \frac{\varepsilon}{k}\right)^k + a\left(\frac{1}{k} - \frac{\varepsilon}{k}\right)^k - \left(1 - \frac{1}{k} - \frac{\varepsilon}{k}\right)^k - a\left(\frac{1}{k} + \frac{\varepsilon}{k}\right)^k \\
&\geq \left(1 - \frac{1}{k} + \frac{\varepsilon}{k}\right)^k + \left(\frac{1}{k} - \frac{\varepsilon}{k}\right)^k - \left(1 - \frac{1}{k} - \frac{\varepsilon}{k}\right)^k - \left(\frac{1}{k} + \frac{\varepsilon}{k}\right)^k \\
&= \mathrm{tr}(\rho_+^k) - \mathrm{tr}(\rho_-^k) \\
&\geq \Omega(\varepsilon),
\end{aligned}$$

where the last inequality is by Equation (9). Using the same reasoning as in the proof of Theorem 4.3, the conclusion immediately follows. □

5 Applications

5.1 Entanglement spectroscopy

Entanglement spectroscopy [JST17] is a powerful tool for analyzing the quantum correlations and topological properties of many-body systems. By studying the *entanglement spectrum*, i.e., the eigenvalues of the reduced density matrix $\rho = \mathrm{tr}_B(|\psi\rangle\langle\psi|)$ obtained from a pure bipartite state $|\psi\rangle$ on systems A and B, one could extract a wealth of information about the many-body systems. For example, the eigenvalues of the reduced state ρ diagnose whether the bipartite state $|\psi\rangle$ is entangled or separable and characterize how strong the entanglement is [HE02]. The entanglement spectrum can be further used to identify topological orders [Fid10, LH08, PTBO10, YQ10], quantum phase transitions [DCLLS12], many-body localizations [SMAP16, YCHM15, YHG+17], and irreversibility in quantum systems [CHM14].

Prior works have reduced the task of entanglement spectroscopy to computing high-order functionals of the reduced density matrix, i.e., $\mathrm{tr}(\rho^k)$ for $k = 2, 3, \ldots, k_{\max}$, which can be done by the Hadamard test [JST17], Two-Copy test [SCC19], qubit reset [YS21], and cyclic shift [QKW24]. Note that $O(k_{\max}^2 \log(k_{\max})/\varepsilon^2)$ samples of ρ are needed in these algorithms to get an estimation

of $\{\mathrm{tr}(\rho^k)\}_{k=2}^{k_{\max}}$ to within additive error ε. It can be seen that this can be further improved to $O(k_{\max}\log(k_{\max})/\varepsilon^2)$, using our simultaneous estimator in Theorem 1.1 for the special case where $\mathcal{O} = I$ is the identity operator.

5.2 Quantum virtual cooling

Quantum virtual cooling [CCL+19] provides a way to estimate properties of thermal states at low temperatures, with a prominent application scenario in the doped Fermi-Hubbard model with ultracold atoms [GB17]. For a thermal state $\rho(T) = e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$ at temperature T with H the Hamiltonian of the system, where $\beta = 1/(k_{\mathrm{B}}T)$ is the inverse temperature and k_{B} is the Boltzmann constant, the expectation of an observable $\mathcal{O}$ with respect to the thermal state $\rho(T/n)$ at the fractional temperature T/n can be expressed as (noted in [CCL+19])

$$\mathrm{tr}(\mathcal{O}\cdot\rho(T/n)) = \frac{\mathrm{tr}(\mathcal{O}\cdot\rho(T)^n)}{\mathrm{tr}(\rho(T)^n)}. \tag{10}$$

Thus, one can estimate $\mathrm{tr}(\mathcal{O}\cdot\rho(T/n))$ from the nonlinear functionals $\mathrm{tr}(\mathcal{O}\cdot\rho(T)^n)$ and $\mathrm{tr}(\rho(T)^n)$ of $\rho(T)$, using $O(n)$ samples of $\rho(T)$.

To draw a tomographic view of the thermal states of a system at different temperatures, one can consider fractional temperatures such as $T/2$, $T/3$, $\ldots$, T/n. Specifically, the expectation $\mathrm{tr}(\mathcal{O}\cdot\rho(T/k))$ at temperature T/k can be obtained from $\mathrm{tr}(\mathcal{O}\cdot\rho(T)^k)$ and $\mathrm{tr}(\rho(T)^k)$ according to Equation (10). This requires to estimate the following values:

$$\begin{array}{cccc} \mathrm{tr}(\mathcal{O}\cdot\rho(T)^2), & \mathrm{tr}(\mathcal{O}\cdot\rho(T)^3), & \ldots, & \mathrm{tr}(\mathcal{O}\cdot\rho(T)^n), \\ \mathrm{tr}(\rho(T)^2), & \mathrm{tr}(\rho(T)^3), & \ldots, & \mathrm{tr}(\rho(T)^n). \end{array}$$

To achieve this, directly employing the approach in [CCL+19] consumes $O(n^2\log(n))$ samples of ρ. In sharp contrast, the simultaneous estimator in Theorem 1.1 allows us to use only $O(n\log(n))$ samples of ρ, yielding a quadratic reduction in resources.

6 Discussion

We present an approach to simultaneously estimating high-order functionals of the form $\mathrm{tr}(\mathcal{O}\rho^k)$ with almost optimal sample complexity. We apply this approach in entanglement spectroscopy and quantum virtual cooling, and extend it to general functionals by polynomial approximation.

An interesting future direction is to extend the simultaneous estimation to more general cases.

1. Multivariate case. The multivariate trace $\mathrm{tr}(\rho_1\rho_2\ldots\rho_m)$ can also be estimated by the cyclic shift [EAO+02], and was recently solved with constant quantum depth [QKW24]. An interesting problem is how to simultaneously estimate $\mathrm{tr}(\rho\sigma)$, $\mathrm{tr}((\rho\sigma)^2)$, $\ldots$, $\mathrm{tr}((\rho\sigma)^k)$, which is related to the estimation of fidelity [QKW24] and multivariate fidelities [NMLW25]. A solution to this problem may inspire new quantum algorithms for estimating quantities involving multiple quantum states, e.g., trace distance [WZ24] and relative entropy [Hay25].

2. Non-integer case. Estimating $\mathrm{tr}(\rho^\alpha)$ for non-integer α is the key step in estimating the Rényi and Tsallis entropies [AISW20, WGL+24, WZL24, WZ25, LW25, CW25]. An interesting problem is how to estimate $\mathrm{tr}(\rho^{\alpha_1}), \mathrm{tr}(\rho^{\alpha_2}), \ldots, \mathrm{tr}(\rho^{\alpha_k})$ for several non-integer $\alpha_1, \alpha_2, \ldots, \alpha_k$. This could lead to fast simultaneous estimation of entropy.

3. Incoherent case. As considered to be near-term friendly, incoherent measurements were investigated for estimating, e.g., $\mathrm{tr}(\rho\sigma)$ and $\mathrm{tr}(\rho^2)$ [ACQ22, CCHL22, ALL22, GHYZ24]. Recently, estimating $\mathrm{tr}(\rho^k)$ for large k with incoherent measurements was studied in [PTTW26]. An interesting question is whether a simultaneous estimation exists in the incoherent case.

4. Query-access case. In addition to sample access, another input model, known as the "purified quantum query access" model [GL20], gives quantum states via state-preparation circuits, as widely employed in the literature (e.g., [GL20, SH21, GHS21, GMF24, GLM+22, WZC+23, WGL+24, GP22, RASW23, WZYW23, WZL24, Wan24, LW25]). However, it is not even clear if we can simultaneously estimate $\mathrm{tr}(\rho^2), \ldots, \mathrm{tr}(\rho^k)$ in this query model. Existing methods based on the generalized SWAP test [EAO+02] and quantum amplitude estimation [BHMT02] estimate $\mathrm{tr}(\rho^j)$ individually to within additive error ε with query complexity $O(j/\varepsilon)$, leading to $O(k^2 \log(k)/\varepsilon)$ for estimating all the k values, which is still worse than the $O(k \log(k)/\varepsilon^2)$ given in this paper when $\varepsilon = \Theta(1)$.

Acknowledgment

The authors would like to thank Mark M. Wilde for helpful discussions, and thank Wang Fang for pointing out a typo in an earlier version of this paper.

The work of Qisheng Wang was supported by the Engineering and Physical Sciences Research Council under Grant EP/X026167/1. The work of Zhicheng Zhang was supported in part by the Australian Research Council under Grant DP250102952 and in part by the Sydney Quantum Academy, NSW, Australia. The work of Zhicheng Zhang was partly carried out during a visit to DIMACS, Rutgers University, USA.

References

[Aar18] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 325–338, 2018. doi:10.1145/3188745.3188802.

[ACQ22] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(1):887, 2022. doi:10.1038/s41467-021-27922-0.

[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.

[AJL09] Dorit Aharonov, Vaughan Jones, and Zeph Landau. A polynomial quantum algorithm for approximating the Jones polynomial. *Algorithmica*, 55(3):395–421, 2009. doi:10.1007/s00453-008-9168-0.

[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.

[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.

[BHMT02] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. In Samuel J. Lomonaco, Jr. and Howard E. Brandt, editors, *Quantum Computation and Information*, volume 305 of *Contemporary Mathematics*, pages 53–74. AMS, 2002. doi:10.1090/conm/305/05215.

[BO21] Costin Bădescu and Ryan O'Donnell. Improved quantum data analysis. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1398–1411, 2021. doi:10.1145/3406325.3451109.

[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.

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

[CCHL22] 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, 2022. doi:10.1109/FOCS52979.2021.00063.

[CCL+19] Jordan Cotler, Soonwon Choi, Alexander Lukin, Hrant Gharibyan, Tarun Grover, M. Eric Tai, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, Adam M. Kaufman, Markus Greiner, Hannes Pichler, and Patrick Hayden. Quantum virtual cooling. *Physical Review X*, 9(3):031013, 2019. doi:10.1103/PhysRevX.9.031013.

[CHM14] Claudio Chamon, Alioscia Hamma, and Eduardo R. Mucciolo. Emergent irreversibility and entanglement spectrum statistics. *Physical Review Letters*, 112(24):240501, 2014. doi:10.1103/PhysRevLett.112.240501.

[CM18] Chris Cade and Ashley Montanaro. The quantum complexity of computing Schatten p-norms. In *Proceedings of the 13th Conference on the Theory of Quantum Computation, Communication and Cryptography*, pages 4:1–4:20, 2018. doi:10.4230/LIPIcs.TQC.2018.4.

[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.

[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.

[DCLLS12] G. De Chiara, L. Lepori, M. Lewenstein, and A. Sanpera. Entanglement spectrum, critical exponents, and order parameters in quantum spin chains. *Physical Review Letters*, 109(23):237208, 2012. doi:10.1103/PhysRevLett.109.237208.

[DTE+25] Zhenyu Du, Yifan Tang, Andreas Elben, Ingo Roth, Jens Eisert, and Zhenhuan Liu. Optimal randomized measurements for a family of non-linear quantum properties. ArXiv e-prints, 2025. arXiv:2505.09206.

[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.

[Fid10] Lukasz Fidkowski. Entanglement spectrum of topological insulators and superconductors. *Physical Review Letters*, 104(13):130502, 2010. doi:10.1103/PhysRevLett.104.130502.

[FL11] Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few Pauli measurements. *Physical Review Letters*, 106(23):230501, 2011. doi:10.1103/PhysRevLett.106.230501.

[GB17] Christian Gross and Immanuel Bloch. Quantum simulations with ultracold atoms in optical lattices. *Science*, 357(6355):995–1001, 2017. doi:10.1126/science.aal3837.

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

[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.

[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.

[GLM+22] András Gilyén, Seth Lloyd, Iman Marvian, Yihui Quek, and Mark M. Wilde. Quantum algorithm for Petz recovery channels and pretty good measurements. *Physical Review Letters*, 128(22):220502, 2022. doi:10.1103/PhysRevLett.128.220502.

[GMF24] Naixu Guo, Kosuke Mitarai, and Keisuke Fujii. Nonlinear transformation of complex amplitudes via quantum singular value transformation. *Physical Review Research*, 6(4):043227, December 2024. doi:10.1103/PhysRevResearch.6.043227.

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

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

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

[HE02] Paweł Horodecki and Artur Ekert. Method for direct detection of quantum entanglement. *Physical Review Letters*, 89(12):127902, 2002. doi:10.1103/PhysRevLett.89.127902.

[Hel67] Carl W. Helstrom. Detection theory and quantum mechanics. *Information and Control*, 10(3):254–291, 1967. doi:10.1016/S0019-9958(67)90302-6.

[HEY+24] Hideaki Hakoshima, Suguru Endo, Kaoru Yamamoto, Yuichiro Matsuzaki, and Nobuyuki Yoshioka. Localized virtual purification. *Physical Review Letters*, 133(8):080601, 2024. doi:10.1103/PhysRevLett.133.080601.

[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.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):337–394, 2020. doi:10.1038/s41567-020-0932-7.

[HKP21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021. doi:10.1103/PhysRevLett.126.190505.

[HMO+21] William J. Huggins, Sam McArdle, Thomas E. O'Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. Virtual distillation for quantum error mitigation. *Physical Review X*, 11(4):041036, 2021. doi:10.1103/PhysRevX.11.041036.

[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.

[Hol73] Alexander S. Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337–394, 1973. doi:10.1016/0047-259X(73)90028-6.

[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.

[Kad52] Richard V. Kadison. A generalized Schwarz inequality and algebraic invariants for operator algebras. *Annals of Mathematics*, 56(3):494–503, 1952. doi:10.2307/1969657.

[KL98] E. Knill and R. Laflamme. Power of one bit of quantum information. *Physical Review Letters*, 81(25):5672, 1998. doi:10.1103/PhysRevLett.81.5672.

[LH08] Hui Li and F. D. M. Haldane. Entanglement spectrum as a generalization of entanglement entropy: identification of topological order in non-abelian fractional quantum Hall effect states. *Physical Review Letters*, 101(1):010504, 2008. doi:10.1103/PhysRevLett.101.010504.

[LLWF23] Jin-Min Liang, Qiao-Qiao Lv, Zhi-Xi Wang, and Shao-Ming Fei. Unified multivariate trace estimation and quantum error mitigation. *Physical Review A*, 107(1):012606, 2023. doi:10.1103/PhysRevA.107.012606.

[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.

[NMLW25] Theshani Nuradha, Hemant K. Mishra, Felix Leditzky, and Mark M. Wilde. Multivariate fidelities. *Journal of Physics A: Mathematical and Theoretical*, 2025. doi:10.1088/1751-8121/adc645.

[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.

[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.

[PTBO10] Frank Pollmann, Ari M. Turner, Erez Berg, and Masaki Oshikawa. Entanglement spectrum of a topological phase in one dimension. *Physical Review B*, 81(6):064439, 2010. doi:10.1103/PhysRevB.81.064439.

[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*, 2026. arXiv:2504.02785.

[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.

[RASW23] Soorya Rethinasamy, Rochisha Agarwal, Kunal Sharma, and Mark M. Wilde. Estimating distinguishability measures on quantum computers. *Physical Review A*, 108(1):012409, 2023. doi:10.1103/PhysRevA.108.012409.

[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.

[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.

[She06] Dan Shepherd. Computation with unitaries and one pure qubit. ArXiv e-prints, 2006. arXiv:quant-ph/0608132.

[SJ08] Peter W. Shor and Stephen P. Jordan. Estimating Jones polynomials is a complete problem for one clean qubit. *Quantum Information and Computation*, 8(8–9):681–714, 2008. doi:10.26421/QIC8.8-9-1.

[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.

[SMAP16] Maksym Serbyn, Alexios A. Michailidis, Dmitry A. Abanin, and Z. Papić. Power-law entanglement spectrum in many-body localized phases. *Physical Review Letters*, 117(16):160601, 2016. doi:10.1103/PhysRevLett.117.160601.

[Wan24] Qisheng Wang. Optimal trace distance and fidelity estimations for pure quantum states. *IEEE Transactions on Information Theory*, 70(12):8791–8805, 2024. doi:10.1109/TIT.2024.3447915.

[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.

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

[WZ24] 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.

[WZ25] 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.

[WZC+23] Qisheng Wang, Zhicheng Zhang, Kean Chen, Ji Guan, Wang Fang, Junyi Liu, and Mingsheng Ying. Quantum algorithm for fidelity estimation. *IEEE Transactions on Information Theory*, 69(1):273–282, 2023. doi:10.1109/TIT.2022.3203985.

[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.

[WZYW23] Youle Wang, Lei Zhang, Zhan Yu, and Xin Wang. Quantum phase processing and its applications in estimating phase and entropies. *Physical Review A*, 108(6):062413, December 2023. doi:10.1103/PhysRevA.108.062413.

[YCHM15] Zhi-Cheng Yang, Claudio Chamon, Alioscia Hamma, and Eduardo R. Mucciolo. Two-component structure in the entanglement spectrum of highly excited states. *Physical Review Letters*, 115(26):267206, 2015. doi:10.1103/PhysRevLett.115.267206.

[YHG+17] Zhi-Cheng Yang, Alioscia Hamma, Salvatore M. Giampaolo, Eduardo R. Mucciolo, and Claudio Chamon. Entanglement complexity in quantum many-body dynamics, thermalization, and localization. *Physical Review B*, 96(2):020408, 2017. doi:10.1103/PhysRevB.96.020408.

[YLLW25] Hongshun Yao, Yingjian Liu, Tengxiang Lin, and Xin Wang. Sample-efficient estimation of nonlinear quantum state functions. ArXiv e-prints, 2025. arXiv:2412.01696v3.

[YQ10] Hong Yao and Xiao-Liang Qi. Entanglement entropy and entanglement spectrum of the Kitaev model. *Physical Review Letters*, 105(8):080501, 2010. doi:10.1103/PhysRevLett.105.080501.

[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.