Nearly Sample-Optimal Estimators for Quantum Rényi and Tsallis Entropies

Kean Chen* and Qisheng Wang†

Abstract

In this paper, we provide estimators for quantum Rényi and Tsallis entropies with nearly optimal sample complexity. Specifically, for order α, dimension d, and additive error ε,

- For $0 < \alpha < 1$, the sample complexity is $O(d^{1+1/\alpha}/\varepsilon^{1/\alpha} + d^{1/\alpha-1}/\varepsilon^2)$ for Rényi entropy and $O(d^{1+1/\alpha}/\varepsilon^{1/\alpha} + d^{2-2\alpha}/\varepsilon^2)$ for Tsallis entropy. In particular, for $0 < \alpha \leq 1/2$, the sample complexity for both entropies is $O(d^{1+1/\alpha}/\varepsilon^{1/\alpha})$.
- For non-integer $\alpha > 1$, the sample complexity is $O(d^2/\varepsilon^{1/\alpha} + d^{1-1/\alpha}/\varepsilon^2)$ for Rényi entropy.

Our upper bounds improve the quantum Rényi entropy estimators due to Acharya, Issa, Shende, and Wagner (2017) and the quantum Tsallis entropy estimators due to Chen, Liu, and Wang (2026), and match the lower bounds recently established by Wang (2026).

*keanchen.gan@gmail.com

†QishengWang1994@gmail.com

Contents

1 Introduction **3**

1.1 Techniques . . . 4

1.2 Related work . . . 6

2 Rényi and Tsallis Entropy with $0 < \alpha < 1$ **6**

2.1 The estimator . . . 7

2.1.1 Tsallis entropy . . . 8

2.1.2 Rényi entropy . . . 8

2.2 Technical lemmas . . . 9

2.2.1 Bures χ^2-divergence bounds for tomography . . . 9

2.2.2 Bias bounds by Bures χ^2-divergence . . . 11

2.2.3 Variance bounds . . . 13

2.2.4 Proof of Lemma 2.3 . . . 16

3 Rényi entropy with $\alpha > 1$ **17**

3.1 The estimator . . . 17

3.2 Main proof . . . 19

3.3 Proof of the one-batch estimates . . . 20

3.3.1 Bias . . . 20

3.3.2 Variance . . . 21

3.4 Technical lemmas . . . 23

3.4.1 Hayashi's POVM . . . 23

3.4.2 Coefficients . . . 24

3.4.3 Schatten norm expansion and moments . . . 25

3.4.4 Some facts . . . 27

References **28**

1 Introduction

Entropy estimation is a fundamental task in information theory and statistics. The estimation of Shannon entropy [Sha48a, Sha48b] has attracted a lot of attention in the literature [Pan03, Pan04, BDKR05, Val11, VV11a, VV11b, VV17, JVHW15, JVHW17, WY16]. In particular, for a d-dimensional discrete distribution P, estimating its Shannon entropy,

$$\mathrm{H}(P) = -\sum_{i=1}^{d} p_i \log(p_i),$$

to within additive error ε is known to have sample complexity $\Theta(\frac{d}{\varepsilon \log(d)} + \frac{\log^2(d)}{\varepsilon^2})$ [JVHW15, WY16]. As generalizations of Shannon entropy, (near-)optimal estimators for Rényi entropy [Rén61] and Tsallis entropy [Tsa88],

$$\mathrm{H}^{\mathrm{R}}_{\alpha}(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{d} p_i^{\alpha}\right), \qquad \mathrm{H}^{\mathrm{T}}_{\alpha}(P) = \frac{1}{1-\alpha} \left(\sum_{i=1}^{d} p_i^{\alpha} - 1\right),$$

have been proposed in [AOST17] and [JVHW15, JVHW17], respectively.

In the quantum world, the estimation of von Neumann entropy [vN27, vN32],

$$\mathrm{S}(\rho) = -\operatorname{tr}(\rho \log(\rho)),$$

has also got extensively studied [BMW16, AISW20, WZ25, GW26]. A nearly sample-optimal estimator for von Neumann entropy was known in [BMW16] with sample complexity $O(\frac{d^2}{\varepsilon} + \frac{\log^2(d)}{\varepsilon^2})$ (cf. [OW17, Theorem 1.27]), where an almost matching lower bound of $\widetilde{\Omega}(\frac{d^2}{\varepsilon} + \frac{1}{\varepsilon^2})$ was recently established in [Wan26].[1] However, the sample complexities of estimating the quantum Rényi entropy and quantum Tsallis entropy,

$$\mathrm{S}^{\mathrm{R}}_{\alpha}(\rho) = \frac{1}{1-\alpha} \log(\mathrm{F}_{\alpha}(\rho)), \qquad \mathrm{S}^{\mathrm{T}}_{\alpha}(\rho) = \frac{1}{1-\alpha} (\mathrm{F}_{\alpha}(\rho) - 1), \qquad \text{where } \mathrm{F}_{\alpha}(\rho) = \operatorname{tr}(\rho^{\alpha}), \tag{1}$$

have not been fully settled.

- **Rényi entropy.**
  - For $0 < \alpha < 1$, the upper bound is $O(\frac{d^{2/\alpha}}{\varepsilon^{2/\alpha}})$ [AISW20], whereas the lower bound $\widetilde{\Omega}(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{1/\alpha-1}}{\varepsilon^2})$ [Wan26].
  - For non-integer $\alpha > 1$, the upper bound is $O(\frac{d^2}{\varepsilon^2})$ [AISW20], whereas the lower bound is $\widetilde{\Omega}(\frac{d^2}{\varepsilon^{1/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2})$ [Wan26].
  - For integer $\alpha \geq 2$, the sample complexity is known to be $\Theta(\frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2})$ [AISW20].
- **Tsallis entropy.**
  - For $0 < \alpha < 1$, the upper bound is $O(\frac{d^{2/\alpha}}{\varepsilon^{2/\alpha}})$ [CLW26], whereas the lower bound is $\widetilde{\Omega}(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{2-2\alpha}}{\varepsilon^2})$ [Wan26].
  - For non-integer $\alpha > 1$, the sample complexity is shown to be $\widetilde{\Theta}(\frac{1}{\varepsilon^{\max\{2/(\alpha-1),2\}}})$ [CW25, Wan26].

[1] A lower bound of $\Omega(d^{2-\gamma})$ was also recently established in [FOW26] for any constant $\gamma > 0$.

– For integer $\alpha \geq 2$, the sample complexity is known to be $\Theta(\frac{1}{\varepsilon^2})$, where the upper bound is by the generalized SWAP test [BCWdW01, EAO+02] and the lower bound was recently established in [CWLY23, GHYZ26, CWYZ26].

In this paper, we close all the remaining gaps between the sample complexity upper and lower bounds by providing near-optimal estimators for (i) quantum Rényi entropy for all non-integer order α and (ii) quantum Tsallis entropy for $0 < \alpha < 1$, leaving only room for polylogarithmic improvements.

Theorem 1.1 (Nearly sample-optimal estimators for quantum Rényi and Tsallis entropies)**.** *Given sample access to an unknown d-dimensional quantum state ρ and the additive error ε,*

- *For $0 < \alpha < 1$, we can estimate the α-Rényi entropy $\mathrm{S}^{\mathrm{R}}_{\alpha}(\rho)$ and Tsallis entropy $\mathrm{S}^{\mathrm{T}}_{\alpha}(\rho)$ using*
$$O\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{1/\alpha-1}}{\varepsilon^2}\right), \qquad O\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{2-2\alpha}}{\varepsilon^2}\right),$$
samples of ρ, respectively. In particular, for $0 < \alpha \leq 1/2$, both estimators have sample complexity
$$O\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}}\right).$$

- *For non-integer $\alpha > 1$, we can estimate the α-Rényi entropy $\mathrm{S}^{\mathrm{R}}_{\alpha}(\rho)$ using*
$$O\left(\frac{d^{2}}{\varepsilon^{1/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2}\right)$$
samples of ρ.

All the sample complexity upper bounds in Theorem 1.1 match the lower bounds in [Wan26] up to polylogarithmic factors in d and $1/\varepsilon$. We summarize the sample complexities of estimating the quantum Rényi and Tsallis entropies in Table 1.

1.1 Techniques

Our estimators are different from the known ones in the literature [AISW20, LW26, CW25, CLW26]. In comparison, the previous estimators build on weak Schur sampling [CHW07] while our estimators build on specific quantum state tomography with new information-theoretic inequalities. Our estimators for $0 < \alpha < 1$ and $\alpha > 1$ are different. So we introduce them below separately.

Some notations and facts. We use $\mathrm{Clip}_{[u,v]}(x) = \min\{\max\{x, u\}, v\}$ to ensure an estimate in a valid range, given the bounds
$$\mathrm{F}_{\alpha}(\rho) = \mathrm{tr}(\rho^{\alpha}) \in \begin{cases} [1, d^{1-\alpha}] & 0 < \alpha < 1, \\ [d^{1-\alpha}, 1] & \alpha > 1. \end{cases} \tag{2}$$

To estimate the Rényi entropy, we estimate $\mathrm{F}_{\alpha}(\rho)$ to a relative error, using the following fact.

Fact 1.2. *For $F, \widehat{F}, \varepsilon > 0$, if $|\widehat{F} - F| \leq (1 - e^{-\varepsilon})F$, then $|\log(F) - \log(\widehat{F})| \leq \varepsilon$.*

Table 1: Sample complexity of estimating quantum Rényi and Tsallis entropies.

<table>
<tr><th>Order</th><th>Rényi Entropy</th><th>Tsallis Entropy</th></tr>
<tr><td>$0 < \alpha \leq 1/2$</td><td colspan="2">$\widetilde{\Theta}\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}}\right)$
This Work↑ and [Wan26]↓</td></tr>
<tr><td>$1/2 < \alpha < 1$</td><td>$\widetilde{\Theta}\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{1/\alpha-1}}{\varepsilon^2}\right)$
This Work↑ and [Wan26]↓</td><td>$\widetilde{\Theta}\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{2-2\alpha}}{\varepsilon^2}\right)$
This Work↑ and [Wan26]↓</td></tr>
<tr><td>$\alpha = 1$
(von Neumann)</td><td colspan="2">$\widetilde{\Theta}\left(\frac{d^2}{\varepsilon} + \frac{1}{\varepsilon^2}\right)$
[BMW16, OW17]↑ and [Wan26]↓</td></tr>
<tr><td>$1 < \alpha < 2$</td><td rowspan="2">$\widetilde{\Theta}\left(\frac{d^2}{\varepsilon^{1/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2}\right)$
This Work↑ and [Wan26]↓</td><td>$\widetilde{\Theta}\left(\frac{1}{\varepsilon^{2/(\alpha-1)}}\right)$
[CW25]↑ and [Wan26]↓</td></tr>
<tr><td>Non-integer $\alpha > 2$</td><td>$\widetilde{\Theta}\left(\frac{1}{\varepsilon^2}\right)$
[CW25]↑↓</td></tr>
<tr><td>Integer $\alpha \geq 2$</td><td>$\Theta\left(\frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2}\right)$
[AISW20]↑↓</td><td>$\Theta\left(\frac{1}{\varepsilon^2}\right)$
[BCWdW01, EAO+02]↑
[CWLY23, GHYZ26, CWYZ26]↓</td></tr>
</table>

↑ References for (near-)optimal upper bounds.
↓ References for (near-)optimal lower bounds.

The case of $0 < \alpha < 1$. Our estimator is based on the new inequality involving the Bures χ^2-divergence (see Lemma 2.6): For d-dimensional density operators ρ and σ with $\sigma \succeq \frac{d}{n} I$,

$$0 \leq (1-\alpha)\operatorname{tr}(\sigma^\alpha) + \alpha \operatorname{tr}\bigl(\rho\sigma^{\alpha-1}\bigr) - \operatorname{tr}(\rho^\alpha) \leq \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma).$$

This suggests that we take $(1-\alpha)\operatorname{tr}(\sigma^\alpha) + \alpha \operatorname{tr}(\rho\sigma^{\alpha-1})$ as an estimate of $\mathrm{F}_\alpha(\rho) = \operatorname{tr}(\rho^\alpha)$. This estimate is good when the Bures χ^2-divergence of ρ to σ is small. To this end, we use quantum state tomography in [PSTW26], which gives an estimate $\widehat{\rho}$ from n samples of ρ; further analysis shows that taking $\sigma = \widehat{\rho} + \frac{2d}{n} I$ yields a small enough Bures χ^2-divergence (see Lemma 2.5):

$$\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) \leq O\left(\frac{d^2}{n}\right).$$

With these, the estimator then consists of two components: (i) estimate $(1-\alpha)\operatorname{tr}(\sigma^\alpha)$, which can be calculated directly from the known σ; and (ii) estimate $\alpha \operatorname{tr}(\rho\sigma^{\alpha-1})$, which can be done by measuring the observable $\sigma^{\alpha-1}$ on ρ. Let m be the number of samples of ρ used in step (ii). The total sample complexity is then simply $n+m$. By choosing n and m appropriately with detailed error analysis, we can yield the sample complexity upper bound presented in Theorem 1.1 for $0 < \alpha < 1$.

The case of $\alpha > 1$. In this case, our estimator is based on a new analysis of Hayashi's pure state estimation algorithm [Hay98] equipped with random purification techniques [TWZ25]. Suppose ρ is an unknown state and $|\psi\rangle$ is a random purification of ρ. Now, we consider a batch of s samples of $|\psi\rangle$. Let $|\widehat{\psi}\rangle$ be the output of Hayashi's state estimation algorithm applied on $|\psi\rangle^{\otimes s}$. We define a one-batch estimate of $F := \mathrm{tr}(\rho^\alpha)$ with batch size s as:

$$Y_s := \mathrm{tr}(\widehat{\rho}^\alpha), \quad \text{where } \widehat{\rho} = \mathrm{tr}_{\mathrm{anc}}(|\widehat{\psi}\rangle\langle\widehat{\psi}|).$$

Central to our proof is the following expansion for the bias (see Lemma 3.4):

$$\mathbb{E}[Y_s] - F = \sum_{j=1}^{k} c_j \mu_j(s) + R(s),$$

where c_j are s-independent coefficients, $R(s)$ is a small remainder with $|R(s)| = O(F \cdot (\frac{d^2}{s})^\alpha)$ and $\mu_j(s)$ is the j-th moment of beta distribution parameterized by s. Then, we repeat the above process for many times for different batch sizes $s_1, \ldots, s_k$. The final estimator for F is defined as $\widehat{F} := \sum_{i=1}^{k} a_i Y_{s_i}$, where $a_1, \ldots, a_k$ are carefully chosen coefficients such that $\sum_{i=1}^{k} a_i = 1$ and $\sum_{i=1}^{k} a_i \mu_j(s_i) = 0$ for all j. This means all the j-th order term for $j \le k$ can be canceled and the bias of $\widehat{F}$ is just a linear combination of the remainders: $\widehat{F} - F = \sum_i a_i R(s_i)$. As $\sum_i |a_i|$ can also be constantly bounded, we can show that the bias $|\widehat{F} - F|$. With a similar but more complicated analysis, the variance of $\widehat{F}$ can also be well bounded (see Lemma 3.5). These produce a good relative error estimate of F, and thus a good additive estimate of the Rényi entropy.

1.2 Related work

The estimation of quantum Rényi and Tsallis entropies has also been studied in other settings. In the incoherent setting, the estimation of quantum Rényi entropy of integer order was studied in [PTTW26]. Given purified quantum query access, quantum Rényi entropy estimation was studied in [SH21, WGL+24, WZL24] and quantum Tsallis entropy estimation was studied in [LW26, Wan25]; for the special case of order $\alpha = 1$ (von Neumann entropy), von Neumann entropy estimation was studied in [GL20, GHS21, WGL+24].

2 Rényi and Tsallis Entropy with $0 < \alpha < 1$

Fix $0 < \alpha < 1$. Both of our estimators for α-Rényi and α-Tsallis entropies first estimate $\mathrm{F}_\alpha(\rho)$ in the same framework, though with different parameter choices. The Tsallis entropy estimator requires additive error, whereas the Rényi estimator requires relative error. The main results of this section are stated as follows.

Theorem 2.1 (Quantum Rényi and Tsallis entropy estimation for $0 < \alpha < 1$)**.** *For $0 < \alpha < 1$, given sample access to an unknown d-dimensional quantum state ρ and the additive error ε, we can estimate the α-Rényi entropy $\mathrm{S}^{\mathrm{R}}_\alpha(\rho)$ and Tsallis entropy $\mathrm{S}^{\mathrm{T}}_\alpha(\rho)$ using*

$$O\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{1/\alpha-1}}{\varepsilon^2}\right), \qquad O\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{2-2\alpha}}{\varepsilon^2}\right),$$

samples of ρ, respectively.

2.1 The estimator

Our estimator is based on the quantum state tomography provided in [PSTW26] which ensures a good precision in Bures χ^2-divergence.

Lemma 2.2 (Bures χ^2-divergence tomography, [PSTW26, Corollary 4.10 and Lemma 6.9])**.** *There is a quantum algorithm* $\mathsf{TomoBures}\chi^2(\rho, d, n)$ *that, on input n samples of an unknown d-dimensional quantum state ρ, outputs a d-dimensional Hermitian matrix*

$$\widehat{\rho} = \sum_{i=1}^{d} \widehat{\lambda}_i |v_i\rangle\langle v_i| \succeq -\frac{d}{n} I, \qquad \widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_d, \tag{3}$$

with $\operatorname{tr}(\widehat{\rho}) = 1$ *such that with probability at least* 0.99*, the following holds:*

1. *Event E: For every unit vector $|w\rangle \in \mathbb{C}^d$,*

$$|\langle w|(\widehat{\rho} - \rho)|w\rangle| \leq C \sqrt{\frac{d}{n} \left(\langle w|\widehat{\rho}|w\rangle + \frac{d}{n} \right)},$$

   where $C > 0$ is a universal constant.

2. *On event E, for every $m \in [d]$,*

$$\sum_{\substack{i,j \in [d] \\ \min\{i,j\} = m}} |\langle v_i|(\widehat{\rho} - \rho)|v_j\rangle|^2 \leq 2C^2 \frac{d}{n} \left(\widehat{\lambda}_m + \frac{d}{n} \right).$$

Let n be the sample size for tomography and let m be the sample size for bias correction. The two batches are disjoint. The formal description of our estimator is given in Algorithm 1.

Algorithm 1 $\mathsf{PowerTrace}(\rho, d, \alpha, n, m)$

Input: Independent samples of an unknown d-dimensional quantum state ρ; $0 < \alpha < 1$; positive integers n, m.
Output: An estimate $\widehat{\mathrm{F}}_\alpha \in [1, d^{1-\alpha}]$ of $\mathrm{F}_\alpha(\rho) = \operatorname{tr}(\rho^\alpha)$.
1: $\widehat{\rho} \leftarrow \mathsf{TomoBures}\chi^2(\rho, d, n)$, using n samples of ρ. ▷ see Lemma 2.2
2: $\sigma \leftarrow \widehat{\rho} + \frac{2d}{n} I$, and decompose $\sigma = \sum_{i=1}^{d} s_i |v_i\rangle\langle v_i|$.
3: **for** $j = 1, \ldots, m$ **do**
4: Measure a sample of ρ in the basis $\{|v_i\rangle\langle v_i|\}_{i=1}^{d}$ and let J_j be the outcome.
5: Set $X_j \leftarrow s_{J_j}^{\alpha-1}$.
6: **end for**
7: $\widetilde{\mathrm{F}}_\alpha \leftarrow (1 - \alpha) \operatorname{tr}(\sigma^\alpha) + \frac{\alpha}{m} \sum_{j=1}^{m} X_j$.
8: **return** $\widehat{\mathrm{F}}_\alpha \leftarrow \mathrm{Clip}_{[1, d^{1-\alpha}]}(\widetilde{\mathrm{F}}_\alpha)$.

Algorithm 1 gives estimates with different precisions depending on the choice of n and m. Below we collect the error results for our purposes.

Lemma 2.3. *Fix $0 < \alpha < 1$. There are constants $C_\alpha^{(0)}, C_\alpha^{(1)} \geq 1$, depending only on α, for which the following holds. Let $d \geq 2$, let ρ be a d-dimensional density operator, and denote* $F := \mathrm{F}_\alpha(\rho) = \operatorname{tr}(\rho^\alpha)$. *Let $\widehat{\mathrm{F}}_\alpha$ be the output of* $\mathsf{PowerTrace}(\rho, d, \alpha, n, m)$ *defined in Algorithm 1, where* $n = \lceil C_\alpha^{(0)} d^{1+1/\alpha} / \varepsilon^{1/\alpha} \rceil$*, and m is to be specified below. Then,* *(i) If* $0 < \alpha \le 1/2$*, then set* $m = \lceil C_\alpha^{(1)} d^{1/\alpha - 1}/\varepsilon^{1/\alpha} \rceil$*, which gives* $\Pr[|\widehat{\mathrm{F}}_\alpha - F| \le \varepsilon] \ge 0.98$.

(ii) If $1/2 < \alpha < 1$*, then set* $m = \lceil C_\alpha^{(1)} d^{2-2\alpha}/\varepsilon^2 \rceil$*, which gives* $\Pr[|\widehat{\mathrm{F}}_\alpha - F| \le \varepsilon] \ge 0.98$.

(iii) If $1/2 < \alpha < 1$*, then, set* $m = \lceil C_\alpha^{(1)} d^{1/\alpha - 1}/\varepsilon^2 \rceil$*, which gives* $\Pr[|\widehat{\mathrm{F}}_\alpha - F| \le \varepsilon F] \ge 0.98$.

For readability, we postpone the proof of Lemma 2.3 in Section 2.2.4.

2.1.1 Tsallis entropy

The formal description of our Tsallis entropy estimator is given in Algorithm 2.

Algorithm 2 $\mathsf{TsallisEntropy}_{<1}(\rho, d, \alpha, \varepsilon)$

Input: Independent samples of an unknown d-dimensional quantum state ρ; $0 < \alpha < 1$; $0 < \varepsilon \le 1$.
Output: An estimate $\widehat{\mathrm{S}}_\alpha^{\mathrm{T}} \in \left[0, (d^{1-\alpha} - 1)/(1-\alpha)\right]$ of $\mathrm{S}_\alpha^{\mathrm{T}}(\rho)$.

1: $\delta \leftarrow (1-\alpha)\varepsilon$.
2: $n \leftarrow \lceil C_\alpha^{(0)} d^{1+1/\alpha} \delta^{-1/\alpha} \rceil$.
3: **if** $0 < \alpha \le 1/2$ **then**
4: $\quad m \leftarrow \lceil C_\alpha^{(1)} d^{1/\alpha - 1} \delta^{-1/\alpha} \rceil$.
5: **else**
6: $\quad m \leftarrow \lceil C_\alpha^{(1)} d^{2-2\alpha} \delta^{-2} \rceil$.
7: **end if**
8: $\widehat{\mathrm{F}}_\alpha \leftarrow \mathsf{PowerTrace}(\rho, d, \alpha, n, m)$.
9: **return** $\widehat{\mathrm{S}}_\alpha^{\mathrm{T}} \leftarrow \dfrac{\widehat{\mathrm{F}}_\alpha - 1}{1-\alpha}$.

Proof of Theorem 2.1 (Tsallis). Let $F := \mathrm{F}_\alpha(\rho)$. The choices of n and m in Algorithm 2 are those prescribed by Lemma 2.3 with $\varepsilon := \delta$. Since $0 < \delta \le 1$, parts (i) and (ii) of that lemma, for $0 < \alpha \le 1/2$ and $1/2 < \alpha < 1$, respectively, give $\Pr[|\widehat{\mathrm{F}}_\alpha - F| \le \delta] \ge 0.98$. On this event, the definition of $\mathrm{S}_\alpha^{\mathrm{T}}$ yields

$$\left|\widehat{\mathrm{S}}_\alpha^{\mathrm{T}} - \mathrm{S}_\alpha^{\mathrm{T}}(\rho)\right| = \frac{1}{1-\alpha}\left|\widehat{\mathrm{F}}_\alpha - F\right| \le \frac{\delta}{1-\alpha} = \varepsilon.$$

Thus the success probability is at least $0.98 > 2/3$.

The total number of samples of ρ is

$$n + m = O_\alpha\!\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{2-2\alpha}}{\varepsilon^2}\right).$$

□

2.1.2 Rényi entropy

The formal description of our Rényi entropy estimator is given in Algorithm 3.

Proof of Theorem 2.1 (Rényi). Let $F := \mathrm{F}_\alpha(\rho)$ and $\delta := 1 - e^{-(1-\alpha)\varepsilon} = \Theta_\alpha(\varepsilon)$. The choices of n and m in Algorithm 3 are those prescribed by Lemma 2.3 with $\varepsilon := \delta$. Since $0 < \delta < 1$, if $0 < \alpha \le 1/2$, part (i) of that lemma gives $|\widehat{\mathrm{F}}_\alpha - F| \le \delta$ with probability at least 0.98. Since $F \ge 1$

Algorithm 3 $\mathsf{RenyiEntropy}_{<1}(\rho, d, \alpha, \varepsilon)$

Input: Independent samples of an unknown d-dimensional quantum state ρ; $0 < \alpha < 1$; $0 < \varepsilon \leq 1$.

Output: An estimate $\widehat{\mathrm{S}}^{\mathrm{R}}_\alpha \in [0, \log d]$ of $\mathrm{S}^{\mathrm{R}}_\alpha(\rho)$.

$\delta \leftarrow 1 - e^{-(1-\alpha)\varepsilon}$.
$n \leftarrow \lceil C^{(0)}_\alpha d^{1+1/\alpha} \delta^{-1/\alpha} \rceil$.
if $0 < \alpha \leq 1/2$ **then**
  $m \leftarrow \lceil C^{(1)}_\alpha d^{1/\alpha - 1} \delta^{-1/\alpha} \rceil$.
else
  $m \leftarrow \lceil C^{(1)}_\alpha d^{1/\alpha - 1} \delta^{-2} \rceil$.
end if
$\widehat{\mathrm{F}}_\alpha \leftarrow \mathsf{PowerTrace}(\rho, d, \alpha, n, m)$.
return $\widehat{\mathrm{S}}^{\mathrm{R}}_\alpha \leftarrow \dfrac{\log(\widehat{\mathrm{F}}_\alpha)}{1-\alpha}$.

by Equation (2), this implies $|\widehat{\mathrm{F}}_\alpha - F| \leq \delta F$. If $1/2 < \alpha < 1$, part (iii) gives the same relative bound directly. Thus, in either regime, $\Pr[|\widehat{\mathrm{F}}_\alpha - F| \leq \delta F] \geq 0.98$. By Fact 1.2, we have

$$\Pr\left[\left|\widehat{\mathrm{S}}^{\mathrm{R}}_\alpha - \mathrm{S}^{\mathrm{R}}_\alpha(\rho)\right| \leq \varepsilon\right] \geq 0.98.$$

The total number of samples of ρ is

$$n + m = O_\alpha\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}} + \frac{d^{1/\alpha-1}}{\varepsilon^2}\right).$$

□

2.2 Technical lemmas

Our error analysis involves the Bures χ^2-divergence.

Definition 2.4 (Bures χ^2-divergence). *Let $\rho, \sigma \in \mathbb{C}^{d\times d}$ be positive semidefinite operators with $\sigma = \sum_{i=1}^d \lambda_i |u_i\rangle\langle u_i|$. If $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, then the Bures χ^2-divergence of ρ to σ is defined by*

$$\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) = \sum_{\substack{i,j\in[d] \\ \lambda_i+\lambda_j>0}} \frac{2}{\lambda_i + \lambda_j} |\langle u_i|(\rho - \sigma)|u_j\rangle|^2.$$

If $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$, define $\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) = +\infty$.

2.2.1 Bures χ^2-divergence bounds for tomography

Lemma 2.5. *Let $\widehat{\rho}$ be the output matrix of $\mathsf{TomoBures}\chi^2(\rho, d, n)$ and set $\sigma = \widehat{\rho} + \frac{2d}{n} I$. Then, $\sigma \succeq \frac{d}{n} I$ and $\mathrm{tr}(\sigma) = 1 + \frac{2d^2}{n}$. Moreover, on event E in Lemma 2.2,*

$$\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) \leq 8(C^2 + 1)\frac{d^2}{n}, \tag{4}$$

where C is the constant in Lemma 2.2.

Proof. By Equation (3) and $\operatorname{tr}(\widehat{\rho}) = 1$, we have

$$\sigma \succeq -\frac{d}{n}I + \frac{2d}{n}I = \frac{d}{n}I, \qquad \operatorname{tr}(\sigma) = \operatorname{tr}(\widehat{\rho}) + \frac{2d}{n}\operatorname{tr}(I) = 1 + \frac{2d^2}{n}.$$

In particular, $\sigma \succ 0$, so the divergence $\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma)$ is finite. Use the eigenbasis in Equation (3),

$$\sigma = \sum_{i=1}^{d} s_i |v_i\rangle\langle v_i|, \qquad s_i = \widehat{\lambda}_i + \frac{2d}{n} \geq \frac{d}{n}.$$

Let $\Delta = \rho - \widehat{\rho}$. Then, $\rho - \sigma = \Delta - \frac{2d}{n}I$, which gives

$$\begin{aligned}
\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) &= \sum_{i=1}^{d}\sum_{j=1}^{d} \frac{2}{s_i + s_j}\left|\langle v_i|\left(\Delta - \frac{2d}{n}I\right)|v_j\rangle\right|^2 \\
&\leq \sum_{i=1}^{d}\sum_{j=1}^{d} \frac{2}{s_i + s_j}\left(2|\langle v_i|\Delta|v_j\rangle|^2 + 2\left|\frac{2d}{n}\langle v_i|v_j\rangle\right|^2\right) \\
&= 4\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{1}{s_i + s_j}|\langle v_i|\Delta|v_j\rangle|^2 + \frac{8d^2}{n^2}\sum_{i=1}^{d}\frac{1}{s_i}.
\end{aligned}$$

Since $s_i \geq d/n$, the second term is then bounded by

$$\frac{8d^2}{n^2}\sum_{i=1}^{d}\frac{1}{s_i} \leq \frac{8d^2}{n}.$$

For the first term, enumerating (i, j) with $\min\{i, j\} = m$ with Lemma 2.2 gives

$$\begin{aligned}
4\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{1}{s_i + s_j}|\langle v_i|\Delta|v_j\rangle|^2 &= 4\sum_{m=1}^{d}\sum_{\substack{i,j\in[d] \\ \min\{i,j\}=m}} \frac{1}{s_i + s_j}|\langle v_i|\Delta|v_j\rangle|^2 \\
&\leq 4\sum_{m=1}^{d}\frac{1}{s_m}\sum_{\substack{i,j\in[d] \\ \min\{i,j\}=m}} |\langle v_i|\Delta|v_j\rangle|^2 \\
&\leq 4\sum_{m=1}^{d}\frac{1}{s_m}\cdot 2C^2\frac{d}{n}\left(\widehat{\lambda}_m + \frac{d}{n}\right) \\
&= 8C^2\frac{d}{n}\sum_{m=1}^{d}\frac{\widehat{\lambda}_m + d/n}{\widehat{\lambda}_m + 2d/n} \\
&\leq 8C^2\frac{d^2}{n}.
\end{aligned}$$

Therefore,

$$\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) \leq 8(C^2 + 1)\frac{d^2}{n}.$$

$\square$

2.2.2 Bias bounds by Bures χ^2-divergence

Lemma 2.6. *Let $0 < \alpha < 1$, let n be a positive integer, let ρ be a d-dimensional density operator, and let $\sigma \succeq \frac{d}{n} I$. Then*

$$0 \le (1-\alpha)\operatorname{tr}(\sigma^\alpha) + \alpha \operatorname{tr}\big(\rho\sigma^{\alpha-1}\big) - \operatorname{tr}(\rho^\alpha) \le \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma). \tag{5}$$

To prove Lemma 2.6, we need the following results in matrix analysis.

Lemma 2.7 ([Bha97, Theorem V.3.3])**.** *Let $f \in C^1(I)$, where I is an interval, and let $A = U\Gamma U^\dagger$ be Hermitian, where U is unitary, $\Gamma = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$, and $\lambda_i \in I$ for every $i \in [d]$. If H is Hermitian, then*

$$\mathrm{D}f(A)[H] \coloneqq \frac{\mathrm{d}}{\mathrm{d}t} f(A+tH)\bigg|_{t=0} = \lim_{t\to 0} \frac{f(A+tH) - f(A)}{t} = U\Big(f^{[1]}(\Gamma) \odot U^\dagger H U\Big)U^\dagger,$$

where $A \odot B$ is the Hadamard element-wise product and $f^{[1]}(\Gamma)$ is a matrix defined by $(f^{[1]}(\Gamma))_{ij} = f^{[1]}(\lambda_i, \lambda_j)$, with

$$f^{[1]}(x,y) = \begin{cases} \dfrac{f(x)-f(y)}{x-y}, & x \neq y, \\ f'(x), & x = y. \end{cases}$$

We need the following lemma, which is a complex-valued version of [LS01, Theorem 3.3].

Lemma 2.8. *Let $f \in C^2((0,\infty))$, let H be Hermitian, and let $A = \operatorname{diag}(a_1, a_2, \ldots, a_d) \succ 0$. Then*

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \operatorname{tr}(f(A+tH))\bigg|_{t=0} = \sum_{i=1}^d \sum_{j=1}^d (f')^{[1]}(a_i, a_j) |H_{ij}|^2.$$

Proof. Let $g(t) \coloneqq \operatorname{tr}(f(A+tH))$. Then, using the fact that $\mathrm{D}(\operatorname{tr}\circ f)(A)[H] = \operatorname{tr}(f'(A)H)$, $g'(t) = \operatorname{tr}(f'(A+tH)H)$. By Lemma 2.7, $(\mathrm{D}f'(A)[H])_{ij} = (f')^{[1]}(a_i, a_j) H_{ij}$. Consequently,

$$\begin{aligned} g''(0) &= \operatorname{tr}\big(\mathrm{D}f'(A)[H]H\big) \\ &= \sum_{i=1}^d \sum_{j=1}^d \big(\mathrm{D}f'(A)[H]\big)_{ij} H_{ji} \\ &= \sum_{i=1}^d \sum_{j=1}^d (f')^{[1]}(a_i, a_j) H_{ij} H_{ji} \\ &= \sum_{i=1}^d \sum_{j=1}^d (f')^{[1]}(a_i, a_j) |H_{ij}|^2, \end{aligned}$$

where the last equality uses $H_{ji} = H_{ij}^*$. ∎

Now we are ready to prove Lemma 2.6.

Proof of Lemma 2.6. Set $\Delta = \rho - \sigma$, $Z_t = (1-t)\sigma + t\rho$, and $g(t) = \operatorname{tr}(Z_t^\alpha)$. Since the map $A \mapsto \operatorname{tr}(A^\alpha)$ is concave,

$$\operatorname{tr}(\rho^\alpha) = g(1) \le g(0) + g'(0) = (1-\alpha)\operatorname{tr}(\sigma^\alpha) + \alpha \operatorname{tr}(\rho\sigma^{\alpha-1}).$$

This proves the first inequality.

For $0 \le t < 1$, $Z_t \succeq (1-t)\frac{d}{n}I$. Let $Z_t = \sum_{i=1}^{d} z_i |v_i\rangle\langle v_i|$ be the spectrum decomposition. Applying Lemma 2.8 to $f(x) = x^\alpha$ in this eigenbasis gives

$$g''(t) = \sum_{i=1}^{d} \sum_{j=1}^{d} (f')^{[1]}(z_i, z_j) |\langle v_i|\Delta|v_j\rangle|^2. \tag{6}$$

Note that for $x > y \ge q > 0$, it holds that

$$-(f')^{[1]}(x, y) \le \alpha \min\{x, y\}^{\alpha-1} \cdot \frac{2}{x+y}. \tag{7}$$

To see this, without loss of generality, we assume that $x \ge y$; denoting $x = wy$ where $w > 1$ and $\beta = 1 - \alpha$, it holds that

$$\begin{aligned}
-(f')^{[1]}(x, y) &= -\alpha \cdot \frac{x^{\alpha-1} - y^{\alpha-1}}{x - y} \\
&= \alpha y^{-\beta-1} \cdot \frac{1 - w^{-\beta}}{w - 1} \\
&\le \alpha y^{-\beta-1} \cdot \frac{2}{w+1} \\
&\le \alpha q^{\alpha-1} \cdot \frac{2}{x+y},
\end{aligned}$$

where the first inequality follows from $w^{-\beta} \ge w^{-1}$. The case $x = y$ follows by continuity.

For a positive definite operator A, define $L_A(H) = AH$ and $R_A(H) = HA$. With the Hilbert–Schmidt inner product $\langle A, B\rangle_{\mathrm{HS}} = \mathrm{tr}(A^\dagger B)$, the definition of $\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma)$ in Definition 2.4 is equivalently $\mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) = 2\langle \Delta, (L_\sigma + R_\sigma)^{-1}(\Delta)\rangle_{\mathrm{HS}}$ (cf. [TKR+10, Definition 1 and Equation (19)]). Combining Equation (7) with Equation (6) gives

$$-g''(t) \le \alpha \left((1-t)\frac{d}{n} \right)^{\alpha-1} \cdot 2\big\langle \Delta, (L_{Z_t} + R_{Z_t})^{-1}(\Delta) \big\rangle_{\mathrm{HS}}.$$

Since $L_{Z_t} + R_{Z_t} \succeq (1-t)(L_\sigma + R_\sigma)$,

$$-g''(t) \le \alpha \left(\frac{d}{n} \right)^{\alpha-1} (1-t)^{\alpha-2} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma). \tag{8}$$

For $0 < u < 1$, Taylor's integral formula gives

$$g(0) + ug'(0) - g(u) = -\int_0^u (u-t) g''(t) \,\mathrm{d}t. \tag{9}$$

We now justify the limit $u \to 1^-$ explicitly. For $0 < u < 1$, extend the integrand to the fixed interval $[0, 1]$ by setting

$$h_u(t) := \begin{cases} (u-t)(-g''(t)), & 0 \le t \le u, \\ 0, & u < t \le 1. \end{cases}$$

Then, Equation (9) becomes

$$g(0) + ug'(0) - g(u) = \int_0^1 h_u(t) \,\mathrm{d}t. \tag{10}$$

Define the limiting function by

$$h(t) := \begin{cases} (1-t)(-g''(t)), & 0 \le t < 1, \\ 0, & t = 1. \end{cases}$$

For every fixed $0 \le t < 1$, once $u > t$ we have $h_u(t) = (u-t)(-g''(t))$, so $h_u(t) \to h(t)$ as $u \to 1^-$. At $t = 1$, the same convergence holds because $h_u(1) = h(1) = 0$.

Moreover, concavity gives $-g''(t) \ge 0$. If $t \le u$, then $0 \le u - t \le 1 - t$, so Equation (8) implies

$$\begin{aligned} 0 \le h_u(t) &\le (1-t)(-g''(t)) \\ &\le \alpha \left(\frac{d}{n}\right)^{\alpha-1} (1-t)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma). \end{aligned}$$

The same bound is immediate when $t > u$, because then $h_u(t) = 0$. Thus the family $\{h_u\}_{0<u<1}$ is dominated independently of u by

$$\Phi(t) := \begin{cases} \alpha \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma)(1-t)^{\alpha-1}, & 0 \le t < 1, \\ 0, & t = 1. \end{cases}$$

This function is integrable on $[0,1]$: indeed, $\alpha - 1 > -1$ and

$$\int_0^1 \Phi(t)\,\mathrm{d}t = \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) < \infty.$$

The dominated convergence theorem therefore gives

$$\lim_{u \to 1^-} \int_0^1 h_u(t)\,\mathrm{d}t = \int_0^1 h(t)\,\mathrm{d}t = \int_0^1 (1-t)(-g''(t))\,\mathrm{d}t, \tag{11}$$

where the value assigned to the last integrand at $t = 1$ is immaterial. On the other hand, when $u \to 1^-$, $Z_u \to \rho$ and the map $A \mapsto \mathrm{tr}(A^\alpha)$ is continuous on the positive semidefinite cone, so $g(u) \to g(1)$; hence the left-hand side of Equation (9) converges to

$$\lim_{u \to 1^-} \big(g(0) + ug'(0) - g(u)\big) = g(0) + g'(0) - g(1). \tag{12}$$

Taking the limit $u \to 1^-$ in Equation (10) and applying Equations (8), (11) and (12) yields

$$\begin{aligned} g(0) + g'(0) - g(1) &\le \alpha \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) \int_0^1 (1-t)^{\alpha-1}\,\mathrm{d}t \\ &= \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma). \end{aligned}$$

This proves Equation (5) by noting that $g(0)+g'(0)-g(1) = (1-\alpha)\,\mathrm{tr}(\sigma^\alpha) + \alpha\,\mathrm{tr}(\rho\sigma^{\alpha-1}) - \mathrm{tr}(\rho^\alpha)$. $\square$

2.2.3 Variance bounds

Lemma 2.9. *Let $\widehat{\rho}$ be the output matrix of* $\mathsf{TomoBures}\chi^2(\rho, d, n)$ *for a d-dimensional quantum state. Let $\sigma = \widehat{\rho} + \frac{2d}{n} I = \sum_{i=1}^d s_i |v_i\rangle\langle v_i|$ and define a random variable X by*

$$X = s_J^{\alpha-1}, \qquad \Pr[J = i] = \langle v_i|\rho|v_i\rangle, \qquad 1 \le i \le d.$$

Then, on event E in Lemma 2.2:

(i) When $0 < \alpha \le 1/2$, then there is a universal constant $C > 0$ such that

$$\mathbb{E}\big[X^2\big] \le Cd\left(\frac{d}{n}\right)^{2\alpha-1}.$$

(ii) When $1/2 < \alpha < 1$, if $n \ge d^2$, then there is a constant $C_\alpha > 0$ depending only on α such that

$$\mathbb{E}\big[X^2\big] \le C_\alpha d^{2-2\alpha}.$$

(iii) When $1/2 < \alpha < 1$, if $n^\alpha \ge d^{\alpha+1}$, then there is a constant $C_\alpha > 0$ depending only on α such that

$$\mathbb{E}\big[X^2\big] \le C_\alpha d^{1/\alpha-1}(\mathrm{tr}(\rho^\alpha))^2.$$

Proof. Set $p_i = \langle v_i|\rho|v_i\rangle$ and $u_i = p_i - s_i = \langle v_i|(\rho-\sigma)|v_i\rangle$. On event E, Lemma 2.5 gives

$$\sum_{i=1}^{d} \frac{u_i^2}{s_i} \le \mathrm{D}_{\chi^2}(\rho\,\|\,\sigma) \le 8(C^2+1)\frac{d^2}{n}. \tag{13}$$

Moreover,

$$\mathbb{E}\big[X^2\big] = \sum_{i=1}^{d} p_i\big(s_i^{\alpha-1}\big)^2 = \sum_{i=1}^{d} s_i^{2\alpha-1} + \sum_{i=1}^{d} u_i s_i^{2\alpha-2}. \tag{14}$$

By the Cauchy–Schwarz inequality, together with Equation (13), we have

$$\left|\sum_{i=1}^{d} u_i s_i^{2\alpha-2}\right| \le \left(\sum_{i=1}^{d}\frac{u_i^2}{s_i}\right)^{1/2}\left(\sum_{i=1}^{d} s_i^{4\alpha-3}\right)^{1/2} \le \sqrt{8(C^2+1)}\frac{d}{\sqrt{n}}\left(\sum_{i=1}^{d} s_i^{4\alpha-3}\right)^{1/2}. \tag{15}$$

Part (i). Suppose $0 < \alpha \le 1/2$. Since $s_i \ge d/n$ and both exponents $2\alpha-1$ and $4\alpha-3$ are nonpositive,

$$\sum_{i=1}^{d} s_i^{2\alpha-1} \le d\left(\frac{d}{n}\right)^{2\alpha-1}, \qquad \sum_{i=1}^{d} s_i^{4\alpha-3} \le d\left(\frac{d}{n}\right)^{4\alpha-3}.$$

Substitution into Equations (14) and (15) gives

$$\mathbb{E}\big[X^2\big] \le \left(\sqrt{8(C^2+1)}+1\right)d\left(\frac{d}{n}\right)^{2\alpha-1}.$$

Part (ii). Now suppose $1/2 < \alpha < 1$ and $n \ge d^2$. Then, Lemma 2.5 gives

$$\sum_{i=1}^{d} s_i = \mathrm{tr}(\sigma) \le 3, \qquad \frac{d}{n} \le \frac{1}{d}.$$

Because $2\alpha - 1 \in (0,1)$, concavity and $\sum_{i=1}^{d} s_i \le 3$ give

$$\sum_{i=1}^{d} s_i^{2\alpha-1} \le d^{1-(2\alpha-1)}\left(\sum_{i=1}^{d} s_i\right)^{2\alpha-1} \le 3^{2\alpha-1}d^{2-2\alpha}. \tag{16}$$

- If $1/2 < \alpha \le 3/4$, then $4\alpha - 3 \le 0$, so $s_i \ge d/n$ implies

$$\sum_{i=1}^{d} s_i^{4\alpha-3} \le d\left(\frac{d}{n}\right)^{4\alpha-3}.$$

  Consequently, Equation (15) gives

$$\begin{aligned}
\left|\sum_{i=1}^{d} u_i s_i^{2\alpha-2}\right| &\le \sqrt{8(C^2+1)}\frac{d}{\sqrt{n}}\left(d\left(\frac{d}{n}\right)^{4\alpha-3}\right)^{1/2} \\
&= \sqrt{8(C^2+1)}d\left(\frac{d}{n}\right)^{2\alpha-1} \\
&\le \sqrt{8(C^2+1)}d^{2-2\alpha},
\end{aligned}$$

  where the last inequality uses $d/n \le 1/d$.

- If $3/4 \le \alpha < 1$, then $4\alpha - 3 \in [0, 1]$. Concavity gives

$$\sum_{i=1}^{d} s_i^{4\alpha-3} \le d^{1-(4\alpha-3)}\left(\sum_{i=1}^{d} s_i\right)^{4\alpha-3} \le 3^{4\alpha-3}d^{4-4\alpha}.$$

  It follows from $d/\sqrt{n} \le 1$ and Equation (15) that

$$\left|\sum_{i=1}^{d} u_i s_i^{2\alpha-2}\right| \le \sqrt{8(C^2+1)}3^{(4\alpha-3)/2}d^{2-2\alpha}.$$

Combining the above two cases with Equations (14) and (16) gives

$$\mathbb{E}\big[X^2\big] \le \Big(3^{2\alpha-1} + \sqrt{8(C^2+1)}\max\Big\{1, 3^{(4\alpha-3)/2}\Big\}\Big)d^{2-2\alpha}.$$

Part (iii). Now suppose $1/2 < \alpha < 1$ and $n^\alpha \ge d^{\alpha+1}$. Let

$$S = \sum_{i=1}^{d} s_i^\alpha, \qquad G = (1-\alpha)S + \alpha\,\mathrm{tr}(\rho\sigma^{\alpha-1}).$$

Both terms in G are nonnegative. Hence $S \le G/(1-\alpha)$; moreover, Lemma 2.6 gives $G \ge \mathrm{tr}(\rho^\alpha) \ge 1$. Since $0 < 2\alpha - 1 < \alpha$, Hölder's inequality gives

$$\sum_{i=1}^{d} s_i^{2\alpha-1} \le d^{1-(2\alpha-1)/\alpha}\left(\sum_{i=1}^{d} s_i^\alpha\right)^{(2\alpha-1)/\alpha}. \tag{17}$$

- If $1/2 < \alpha < 3/4$, then $s_i^{4\alpha-3} \le (d/n)^{2\alpha-2}s_i^{2\alpha-1}$. By Equations (15) and (17), we have

$$\begin{aligned}
\left|\sum_{i=1}^{d} u_i s_i^{2\alpha-2}\right| &\le \sqrt{8(C^2+1)}\,d^{1/(2\alpha)}\left(\frac{d}{n}\right)^{\alpha-1/2}S^{(2\alpha-1)/(2\alpha)} \\
&\le \sqrt{8(C^2+1)}\,d^{1/\alpha-1}S^{(2\alpha-1)/(2\alpha)} \\
&\le \sqrt{8(C^2+1)}\,(1-\alpha)^{-(2\alpha-1)/(2\alpha)}d^{1/\alpha-1}G^2.
\end{aligned}$$

  The last inequality again follows from $S \le G/(1-\alpha)$ and $G \ge 1$.

- If $3/4 \le \alpha < 1$, put $r = 4\alpha - 3$. For $r > 0$, Hölder's inequality gives $\sum_i s_i^r \le d^{1-r/\alpha} S^{r/\alpha}$; for $r = 0$, the same formula is the identity $\sum_i s_i^0 = d$. Substituting this bound into Equation (15) gives

$$\begin{aligned}\left|\sum_{i=1}^{d} u_i s_i^{2\alpha-2}\right| &\le \sqrt{8(C^2+1)}\, d^{1-r/(2\alpha)} \left(\frac{d}{n}\right)^{1/2} S^{r/(2\alpha)} \\ &\le \sqrt{8(C^2+1)}\, d^{1/\alpha-1} S^{r/(2\alpha)} \\ &\le \sqrt{8(C^2+1)}\, (1-\alpha)^{-r/(2\alpha)} d^{1/\alpha-1} G^2.\end{aligned}$$

Here the second inequality uses $d/n \le d^{-1/\alpha}$ and $r+1 = 4\alpha - 2$, while the last inequality uses $S \le G/(1-\alpha)$ and $G^{r/(2\alpha)} \le G^2$.

Combining the above two cases with Equations (14) and (16) gives

$$\mathbb{E}\left[X^2\right] \le \left(3^{2\alpha-1} + \sqrt{8(C^2+1)}\max\left\{1, 3^{(4\alpha-3)/2}\right\}\right) d^{2-2\alpha}.$$

Since $4\alpha - 3 \le 2\alpha - 1$, the above two cases gives

$$\left|\sum_{i=1}^{d} u_i s_i^{2\alpha-2}\right| \le \sqrt{8(C^2+1)}\,(1-\alpha)^{-(2\alpha-1)/(2\alpha)} d^{1/\alpha-1} G^2.$$

Together with Equations (14) and (17), this gives

$$\mathbb{E}\left[X^2\right] \le \left((1-\alpha)^{-(2\alpha-1)/\alpha} + \sqrt{8(C^2+1)}\,(1-\alpha)^{-(2\alpha-1)/(2\alpha)}\right) d^{1/\alpha-1} G^2. \tag{18}$$

Moreover, Lemmas 2.5 and 2.6 and $n^\alpha \ge d^{\alpha+1}$ give

$$0 \le G - \operatorname{tr}(\rho^\alpha) \le \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) \le 8(C^2+1)\frac{d^{\alpha+1}}{n^\alpha} \le 8(C^2+1).$$

Since $\mathrm{F}_\alpha(\rho) \ge 1$, it follows that $G \le (8(C^2+1)+1)\operatorname{tr}(\rho^\alpha)$. Substituting this into Equation (18) gives $\mathbb{E}[X^2] \le C_\alpha d^{1/\alpha-1}(\operatorname{tr}(\rho^\alpha))^2$ with

$$C_\alpha = \left(8(C^2+1)+1\right)^2 \left((1-\alpha)^{-(2\alpha-1)/\alpha} + \sqrt{8(C^2+1)}\,(1-\alpha)^{-(2\alpha-1)/(2\alpha)}\right).$$

□

2.2.4 Proof of Lemma 2.3

Using the above technical lemmas, we prove Lemma 2.3 as follows.

Proof of Lemma 2.3. Let $\widehat{\rho}$ be the tomography output of $\mathsf{TomoBures}\chi^2(\rho, d, n)$ and write $\sigma = \widehat{\rho} + \frac{2d}{n} I = \sum_{i=1}^d s_i |v_i\rangle\langle v_i|$, as in Algorithm 1. Note that X_j for $1 \le j \le m$ are independent and identically distributed, with

$$\Pr[J_j = i \mid \sigma] = \langle v_i|\rho|v_i\rangle, \qquad X_j = s_{J_j}^{\alpha-1}.$$

Consequently,

$$\mathbb{E}\left[\widetilde{\mathrm{F}}_\alpha\right] = (1-\alpha)\operatorname{tr}(\sigma^\alpha) + \alpha \operatorname{tr}(\rho\sigma^{\alpha-1}). \tag{19}$$

Let E be the event in Lemma 2.2 with $\Pr[E] \geq 0.99$. On event E, Lemmas 2.5 and 2.6 and Equation (19) give

$$0 \leq \mathbb{E}\Big[\widetilde{\mathrm{F}}_\alpha\Big] - F \leq \left(\frac{d}{n}\right)^{\alpha-1} \mathrm{D}_{\chi^2}(\rho \,\|\, \sigma) = O_\alpha\!\left(\frac{d^{\alpha+1}}{n^\alpha}\right) = O_\alpha(\varepsilon). \tag{20}$$

On the other hand,

$$\mathbf{Var}\Big[\widetilde{\mathrm{F}}_\alpha\Big] = \frac{\alpha^2}{m}\mathbf{Var}[X_1] \leq \frac{\alpha^2}{m}\mathbb{E}\big[X_1^2\big].$$

By Chebyshev's inequality,

$$\Pr\Big[\Big|\widetilde{\mathrm{F}}_\alpha - \mathbb{E}\Big[\widetilde{\mathrm{F}}_\alpha\Big]\Big| \leq t\Big] \geq 1 - \frac{\mathbf{Var}[\widetilde{\mathrm{F}}_\alpha]}{t^2} \geq 1 - \frac{\alpha^2}{mt^2}\mathbb{E}\big[X_1^2\big]. \tag{21}$$

Part (i). Suppose first that $0 < \alpha \leq 1/2$. Part (i) of Lemma 2.9 gives that on E, $\mathbb{E}[X_1^2] \leq O(d^{2\alpha}n^{1-2\alpha})$. By Equation (21), with appropriate choice of $C_\alpha^{(0)}$ and $C_\alpha^{(1)}$, we have $\Pr[|\widetilde{\mathrm{F}}_\alpha - \mathbb{E}[\widetilde{\mathrm{F}}_\alpha]| \leq \varepsilon/2] \geq 0.99$. Together with Equation (20), this yields $\Pr[|\widetilde{\mathrm{F}}_\alpha - F| \leq \varepsilon] \geq 0.98$.

Part (ii). Suppose that $1/2 < \alpha < 1$. The choice of n satisfies $n \geq d^{1+1/\alpha} \geq d^2$. Hence part (ii) of Lemma 2.9 gives that on E, $\mathbb{E}[X_1^2] \leq O_\alpha(d^{2-2\alpha})$. By Equation (21), with $t = \varepsilon/2$ and $C_\alpha^{(1)}$ sufficiently large, we have on E, $\Pr[|\widetilde{\mathrm{F}}_\alpha - \mathbb{E}[\widetilde{\mathrm{F}}_\alpha]| \leq \varepsilon/2] \geq 0.99$. Together with Equation (20), this gives on E, $\Pr[|\widetilde{\mathrm{F}}_\alpha - F| \leq \varepsilon] \geq 0.98$.

Part (iii). Suppose that $1/2 < \alpha < 1$. The choice of n satisfies $n^\alpha \geq d^{\alpha+1}$. Hence part (iii) of Lemma 2.9 and Equation (2) give that on E, $\mathbb{E}\big[X_1^2\big] \leq C_\alpha d^{1/\alpha-1}F^2$. By Equation (21), with $t = \varepsilon F/2$ and $C_\alpha^{(1)}$ sufficiently large, we have on E, $\Pr[|\widetilde{\mathrm{F}}_\alpha - \mathbb{E}[\widetilde{\mathrm{F}}_\alpha]| \leq \varepsilon F/2] \geq 0.99$. Since Equation (20) gives $0 \leq \mathbb{E}[\widetilde{\mathrm{F}}_\alpha] - F \leq \varepsilon/2 \leq \varepsilon F/2$, we obtain on E, $\Pr[|\widetilde{\mathrm{F}}_\alpha - F| \leq \varepsilon F] \geq 0.98$. ☐

3 Rényi entropy with $\alpha > 1$

In this section, we prove the following result.

Theorem 3.1 (Rényi entropy estimation for $\alpha > 1$). *For non-integer $\alpha > 1$, given sample access to an unknown d-dimensional quantum state ρ and the additive error ε, we can estimate the α-Rényi entropy $\mathrm{S}_\alpha^{\mathrm{R}}(\rho)$ using*

$$O\!\left(\frac{d^2}{\varepsilon^{1/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2}\right)$$

samples of ρ.

3.1 The estimator

First, our estimator requires the following subroutines.

Lemma 3.2 (Random purification, [TWZ25, Lemma 2.11]). *For every integer $s \geq 1$, there is a channel $\Lambda_{\mathrm{pur}}^{(s)}$ such that for every d-dimensional quantum state ρ,*

$$\Lambda_{\mathrm{pur}}^{(s)}(\rho^{\otimes s}) = \mathop{\mathbb{E}}_{|\psi_\rho\rangle}\big[|\psi_\rho\rangle\!\langle\psi_\rho|^{\otimes s}\big],$$

where $|\psi_\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ is sampled uniformly from the space of purifications of ρ, meaning that $\mathrm{tr}_{\mathrm{anc}}(|\psi_\rho\rangle\!\langle\psi_\rho|) = \rho$.

Definition 3.3 (Hayashi's POVM [Hay98])**.** *For a pure state $|\psi\rangle \in \mathbb{C}^{d^2}$, Hayashi's covariant POVM on* $\mathrm{Sym}^s(\mathbb{C}^{d^2})$ *is*

$$\dim\big(\mathrm{Sym}^s(\mathbb{C}^{d^2})\big)\,|v\rangle\langle v|^{\otimes s}\,\mathrm{d}v = \binom{d^2+s-1}{s}|v\rangle\langle v|^{\otimes s}\mathrm{d}v, \tag{22}$$

where $\mathrm{d}v$ *is Haar probability measure. Note that on input $|\psi\rangle^{\otimes s}$, the outcome density is proportional to $|\langle v|\psi\rangle|^{2s}$.*

Now we are ready to define our estimator. Fix an order $\alpha > 1$ and an accuracy $0 < \varepsilon \leq 1$. Set

$$\theta := 1 - e^{-(\alpha-1)\varepsilon}, \qquad k := \lceil \alpha \rceil - 1. \tag{23}$$

Fix an α-dependent constant $K_\alpha \geq 1$ (defined later in Equation (31)) and define

$$m := \left\lceil K_\alpha \left(\frac{d^2}{\theta^{1/\alpha}} + \frac{d^{1-1/\alpha}}{\theta^2} \right) \right\rceil, \qquad s_\ell := 2^\ell m \quad (0 \leq \ell \leq k). \tag{24}$$

Because $\theta \leq 1$, every s_ℓ is at least d^2. For integers $j \geq 0$ and $s \geq 1$, put

$$\mu_0(s) := 1, \qquad \mu_j(s) := \prod_{r=0}^{j-1} \frac{d^2-1+r}{d^2+s+r} \quad (j \geq 1). \tag{25}$$

Let $a_0, \ldots, a_k$ be the unique solution of

$$\sum_{\ell=0}^{k} a_\ell = 1, \qquad \sum_{\ell=0}^{k} a_\ell \mu_j(s_\ell) = 0 \quad (1 \leq j \leq k). \tag{26}$$

The existence, uniqueness, and uniform boundedness of these coefficients are proved in Lemma 3.7.

Algorithm 4 $\mathsf{RenyiEntropy}_{>1}(\rho, d, \alpha, \varepsilon)$

Input: Independent samples of an unknown d-dimensional quantum state ρ; $\alpha > 1$; $0 < \varepsilon \leq 1$.

Compute $\theta, k, m, s_0, \ldots, s_k$ from Equations (23) and (24).
Compute $\mu_j(s_\ell)$ from Equation (25) and solve Equation (26) for $a_0, \ldots, a_k$.
for $\ell = 0, \ldots, k$ **do**
    Reserve a fresh batch of s_ℓ samples of ρ.
    Apply the random-purification channel $\Lambda_{\mathrm{pur}}^{(s_\ell)}$ to this batch. ▷ see Lemma 3.2
    Apply the covariant pure-state POVM to the channel output and denote its outcome by $|v_\ell\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. ▷ see Definition 3.3
    Set $\widehat{\rho}_\ell \leftarrow \mathrm{tr}_{\mathrm{anc}}|v_\ell\rangle\langle v_\ell|$ and $Y_\ell \leftarrow \mathrm{tr}(\widehat{\rho}_\ell^{\alpha})$.
end for
Set $\widetilde{\mathrm{F}}_\alpha \leftarrow \sum_{\ell=0}^{k} a_\ell Y_\ell$.
Set $\widehat{\mathrm{F}}_\alpha \leftarrow \mathrm{Clip}_{[d^{1-\alpha},1]}(\widetilde{\mathrm{F}}_\alpha)$.
return $\widehat{\mathrm{S}}_\alpha^{\mathrm{R}} \leftarrow \dfrac{\log(\widehat{\mathrm{F}}_\alpha)}{1-\alpha}$.

3.2 Main proof

For this section only, abbreviate $F := \mathrm{F}_\alpha(\rho)$. To prove Theorem 3.1, we need the following two lemmas.

Lemma 3.4 (One-batch estimates)**.** *Let $s \geq d^2$, and let Y_s be the power estimate obtained from one batch of size s in Algorithm 4. There are coefficients $c_1(\rho), \ldots, c_k(\rho)$ independent of s, such that*

$$\mathbb{E}[Y_s] = F + \sum_{j=1}^{k} c_j(\rho)\mu_j(s) + R_s, \qquad |c_j(\rho)| \leq O_\alpha(F), \qquad |R_s| \leq O_\alpha\!\left(F \cdot \left(\frac{d^2}{s}\right)^{\alpha}\right). \tag{27}$$

Moreover,

$$\mathbf{Var}[Y_s] \leq O_\alpha\!\left(F^2 \cdot \left[\frac{d^{1-1/\alpha}}{s} + \left(\frac{d^2}{s}\right)^{2\alpha}\right]\right). \tag{28}$$

The proof of Lemma 3.4 is given in Section 3.3.

Lemma 3.5 (Bias and variance)**.** *The raw estimator produced by Algorithm 4 satisfies: there exists constants $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$ depending only on α, such that*

$$\left|\mathbb{E}\left[\widetilde{\mathrm{F}}_\alpha\right] - F\right| \leq C_\alpha^{(1)} F \left(\frac{d^2}{m}\right)^{\alpha}, \tag{29}$$

$$\mathbf{Var}[\widetilde{\mathrm{F}}_\alpha] \leq C_\alpha^{(2)} F^2 \left[\frac{d^{1-1/\alpha}}{m} + \left(\frac{d^2}{m}\right)^{2\alpha}\right]. \tag{30}$$

Proof. By Lemma 3.4,

$$\mathbb{E}[Y_\ell] = F + \sum_{j=1}^{k} c_j(\rho)\mu_j(s_\ell) + R_{s_\ell}.$$

The coefficients $c_j(\rho)$ are common to all batches. Hence Equation (26) cancels the displayed polynomial terms and gives

$$\mathbb{E}\left[\widetilde{\mathrm{F}}_\alpha\right] - F = \sum_{\ell=0}^{k} a_\ell R_{s_\ell}.$$

Using $s_\ell \geq m$ with Equation (27), and the coefficient bound in Lemma 3.7, we can prove Equation (29).

The batches use disjoint input samples and separate channels and measurements, so $Y_0, \ldots, Y_k$ are independent. Therefore

$$\mathbf{Var}\left[\widetilde{\mathrm{F}}_\alpha\right] = \sum_{\ell=0}^{k} a_\ell^2\, \mathbf{Var}[Y_\ell].$$

Now apply Equation (28), $s_\ell \geq m$, and Lemma 3.7. □

Then, we are able to prove Theorem 3.1.

Proof of Theorem 3.1. Let C_α be a constant dominating the constants in Lemma 3.5. Choose the fixed tuning constant K_α large enough that

$$C_\alpha K_\alpha^{-\alpha} \leq \frac{1}{4}, \qquad C_\alpha\big(K_\alpha^{-1} + K_\alpha^{-2\alpha}\big) \leq \frac{1}{100}. \tag{31}$$

This choice depends only on α. From Equation (24),

$$\left(\frac{d^2}{m}\right)^{\alpha} \le K_\alpha^{-\alpha}\theta, \qquad \frac{d^{1-1/\alpha}}{m} \le K_\alpha^{-1}\theta^2, \qquad \left(\frac{d^2}{m}\right)^{2\alpha} \le K_\alpha^{-2\alpha}\theta^2.$$

Consequently, Lemma 3.5 yields

$$\left|\mathbb{E}\left[\widetilde{\mathrm{F}}_\alpha\right] - F\right| \le \frac{1}{4}\theta F, \qquad \mathbf{Var}[\widetilde{\mathrm{F}}_\alpha] \le \frac{1}{100}\theta^2 F^2. \tag{32}$$

Chebyshev's inequality gives $\mathbb{P}[|\widetilde{\mathrm{F}}_\alpha - \mathbb{E}[\widetilde{\mathrm{F}}_\alpha]| > \frac{1}{2}\theta F] \le \frac{1}{25}$. On the complementary event, $|\widetilde{\mathrm{F}}_\alpha - F| \le \frac{3}{4}\theta F \le \theta F$. Since $F \in [d^{1-\alpha}, 1]$, the clip operation into this interval does not increase the error. Thus $|\widehat{\mathrm{F}}_\alpha - F| \le \theta F$. The projected estimate is positive, and Fact 1.2, with $\gamma = (\alpha-1)\varepsilon$ and $\theta = 1 - e^{-\gamma}$, implies $|\log \widehat{\mathrm{F}}_\alpha - \log F| \le (\alpha-1)\varepsilon$. Dividing by $\alpha - 1$ proves the asserted entropy error. The success probability is at least $24/25$, and hence at least $2/3$.

Finally, note that $\sum_{\ell=0}^{k} s_\ell = (2^{k+1}-1)m = O(m)$, and k depends only on α. Moreover, $\theta = \Theta(\varepsilon)$. Therefore, the final sample complexity is

$$O(m) = O\left(\frac{d^2}{\varepsilon^{1/\alpha}} + \frac{d^{1-1/\alpha}}{\varepsilon^2}\right).$$

□

3.3 Proof of the one-batch estimates

Now, we give the proof of Lemma 3.4. Fix a purification $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ of ρ, we identify it as $|M\rangle\rangle$ for a coefficient matrix $M \in \mathbb{C}^{d\times d}$. Then

$$\|M\|_2 = \|M\|_{\mathrm{F}} = 1, \qquad \rho = MM^\dagger, \qquad F = \mathrm{F}_\alpha(\rho) = \|M\|_{2\alpha}^{2\alpha}.$$

By Lemma 3.6, the coefficient matrix of the outcome of the covariant pure-state POVM (see line 6 of Algorithm 4) is $\widehat{M}_T = \sqrt{1-T}\,M + \sqrt{T}\,G$, where $T \sim \mathrm{Beta}(d^2-1, s+1)$ and G is uniform on the unit sphere of $M^\perp$, independently of T. The one-batch statistic is $Y_s = \|\widehat{M}_T\|_{2\alpha}^{2\alpha}$.

3.3.1 Bias

For $0 \le t \le 1/2$, set $u := \sqrt{t/(1-t)} \le 1$. By homogeneity,

$$\mathop{\mathbb{E}}_G[Y_s \mid T = t] = \mathop{\mathbb{E}}_G\left[\|\sqrt{1-t}M + \sqrt{t}G\|_{2\alpha}^{2\alpha}\right] = (1-t)^\alpha \mathop{\mathbb{E}}_G\left[\|M + uG\|_{2\alpha}^{2\alpha}\right]. \tag{33}$$

Since $\|M\|_p \le \|M\|_{\mathrm{F}} = 1$ and $\|G\|_p \le \|G\|_{\mathrm{F}} = 1$, Lemma 3.8 applies. The law of G is invariant under $G \mapsto -G$, so every odd multilinear term has zero expectation. The largest even integer not exceeding $\lceil 2\alpha \rceil - 1$ is $2k$ (see Equation (23)). Thus, with $b_j := \mathbb{E}_G[L_{2j,M}(G,\ldots,G)]$,

$$\mathop{\mathbb{E}}_G\left[\|M + uG\|_{2\alpha}^{2\alpha}\right] = F + \sum_{j=1}^{k} b_j u^{2j} + \mathcal{E}(u). \tag{34}$$

The multilinear bound in Lemma 3.8 and Lemma 3.9 give

$$|b_j| \le O_\alpha\left(\|M\|_{2\alpha}^{2\alpha-2j}\,\mathbb{E}\left[\|G\|_{2\alpha}^{2j}\right]\right) \le O_\alpha\left(F^{1-j/\alpha} d^{j(1/\alpha-1)}\right) \le O_\alpha(F),$$

where the last step uses $d^{1/\alpha-1} \le F^{1/\alpha}$. Likewise,

$$|\mathcal{E}(u)| \le O_\alpha\big(u^{2\alpha}\,\mathbb{E}\big[\|G\|_{2\alpha}^{2\alpha}\big]\big) \le O_\alpha\big(Fu^{2\alpha}\big)\,. \tag{35}$$

Substituting Equations (34) and (35) into Equation (33) yields, for $0 \le t \le 1/2$,

$$\mathop{\mathbb{E}}_{G}[Y_s \mid T=t] = F(1-t)^\alpha + \sum_{j=1}^{k} b_j t^j (1-t)^{\alpha-j} + O_\alpha(Ft^\alpha). \tag{36}$$

For each $0 \le j \le k$, apply Lemma 3.10 with $\beta = \alpha - j$. Since $\lceil \alpha - j \rceil - 1 = k - j$, multiplying the resulting polynomial by t^j produces a polynomial of degree at most k and an error of order t^α. Hence there are coefficients $c_1(\rho), \ldots, c_k(\rho)$ with $|c_j(\rho)| \le O_\alpha(F)$ such that

$$\mathop{\mathbb{E}}_{G}[Y_s \mid T=t] = F + \sum_{j=1}^{k} c_j(\rho) t^j + r(t), \qquad |r(t)| \le O_\alpha(Ft^\alpha) \tag{37}$$

for $0 \le t \le 1/2$.

The same bound extends to $1/2 \le t \le 1$. Indeed, the Schatten triangle inequality and Lemma 3.9 imply

$$\mathop{\mathbb{E}}_{G}[Y_s \mid T=t] = \mathop{\mathbb{E}}_{G}\Big[\|\sqrt{1-t}M + \sqrt{t}G\|_{2\alpha}^{2\alpha}\Big] \le O_\alpha\big(F + \mathbb{E}\big[\|G\|_{2\alpha}^{2\alpha}\big]\big) \le O_\alpha(F).$$

The polynomial in Equation (37) is also bounded by $O_\alpha(F)$ on $[0,1]$. For $1/2 \le t \le 1$, let

$$r(t) := \mathop{\mathbb{E}}_{G}[Y_s \mid T=t] - \left(F + \sum_{j=1}^{k} c_j(\rho) t^j\right).$$

Thus $|r(t)| \le O_\alpha(F) \le O_\alpha(Ft^\alpha)$ since $t^\alpha \ge 2^{-\alpha}$ on $[1/2, 1]$. Therefore Equation (37) holds for all $0 \le t \le 1$.

Averaging Equation (37) over T and using Lemma 3.6 proves Equation (27). Finally, replacing M by MU for a unitary U on the purifying register sends G to GU, preserves the Frobenius hyperplane and all Schatten norms, and therefore leaves the conditional law and the coefficients unchanged. Thus the coefficients depend only on ρ, and the same expansion holds after the random-purification channel.

3.3.2 Variance

For fixed t, write

$$h_t(G) := \|\sqrt{1-t}M + \sqrt{t}G\|_{2\alpha}^{2\alpha}, \qquad \phi(t) := \mathop{\mathbb{E}}_{G}[h_t(G)].$$

The law of total variance gives

$$\mathbf{Var}[Y_s] = \mathop{\mathbb{E}}_{T}\Big[\mathop{\mathbf{Var}}_{G}[h_T(G) \mid T]\Big] + \mathop{\mathbf{Var}}_{T}[\phi(T)]. \tag{38}$$

For the first term, apply Lemma 3.13 with $H = M^\perp$, viewed as a real Hilbert space with inner product $\langle A, B\rangle_{\mathbb{R}} := \operatorname{Re}\operatorname{tr}(A^\dagger B)$. The complex dimension of $M^\perp$ is d^2-1, so its real dimension is $m = 2d^2 - 2$. Consequently, its unit sphere has real dimension $m - 1 = 2d^2 - 3$. For fixed t, consider the ambient extension

$$\widetilde{h}_t(Z) := \|\sqrt{1-t}M + \sqrt{t}Z\|_{2\alpha}^{2\alpha}, \qquad Z \in M^\perp,$$

so that h_t is the restriction of $\widetilde{h}_t$ to the unit sphere of $M^\perp$. By Equation (53),

$$\operatorname*{\mathbf{Var}}_G[h_t(G)] \le \frac{1}{2d^2-3}\mathop{\mathbb{E}}_G\left[\left\|\nabla_{M^\perp}\widetilde{h}_t(G)\right\|_{\mathrm{F}}^2\right].$$

The gradient in $M^\perp$ is the orthogonal projection of the full matrix gradient. We set

$$\widehat{M}_t := \sqrt{1-t}M + \sqrt{t}G.$$

Thus, by the chain rule,

$$\left\|\nabla_{M^\perp}\widetilde{h}_t(G)\right\|_{\mathrm{F}} \le \sqrt{t}\,\left\|\nabla\|X\|_{2\alpha}^{2\alpha}|_{X=\widehat{M}_t}\right\|_{\mathrm{F}}.$$

Using the Frobenius-gradient identity

$$\left\|\nabla_X\|X\|_{2\alpha}^{2\alpha}\right\|_{\mathrm{F}}^2 = (2\alpha)^2\operatorname{tr}\left[(XX^\dagger)^{2\alpha-1}\right],$$

we obtain

$$\begin{aligned}
\operatorname*{\mathbf{Var}}_G[h_t(G)] &\le \frac{t}{2d^2-3}\mathop{\mathbb{E}}_G\left[\left\|\nabla_X\|X\|_{2\alpha}^{2\alpha}|_{X=\widehat{M}_t}\right\|_{\mathrm{F}}^2\right]\\
&= \frac{(2\alpha)^2 t}{2d^2-3}\mathop{\mathbb{E}}_G\left[\operatorname{tr}\left[(\widehat{M}_t\widehat{M}_t^\dagger)^{2\alpha-1}\right]\right]\\
&\le O_\alpha\left(\frac{t}{d^2}\mathop{\mathbb{E}}_G\left[\operatorname{tr}\left[(\widehat{M}_t\widehat{M}_t^\dagger)^{2\alpha-1}\right]\right]\right).
\end{aligned}$$

The Schatten triangle inequality, followed by Lemma 3.9 with $q = 4\alpha-2$, yields

$$\begin{aligned}
\mathop{\mathbb{E}}_G\left[\operatorname{tr}\left[(\widehat{M}_t\widehat{M}_t^\dagger)^{2\alpha-1}\right]\right] &= \mathop{\mathbb{E}}_G\left[\|\sqrt{1-t}M+\sqrt{t}G\|_{4\alpha-2}^{4\alpha-2}\right]\\
&\le O_\alpha\left((1-t)^{2\alpha-1}\mathop{\mathbb{E}}_G\left[\|M\|_{4\alpha-2}^{4\alpha-2}\right] + t^{2\alpha-1}\mathop{\mathbb{E}}_G\left[\|G\|_{4\alpha-2}^{4\alpha-2}\right]\right)\\
&= O_\alpha\left(\operatorname{tr}(\rho^{2\alpha-1}) + t^{2\alpha-1}d^{2-2\alpha}\right).
\end{aligned}$$

Average over T and use Lemma 3.6. Since $\mathbb{E}[T]/d^2 \le O(1/s)$, this gives

$$\begin{aligned}
\mathop{\mathbb{E}}_T\left[\operatorname*{\mathbf{Var}}_G[h_T(G)\mid T]\right] &\le O_\alpha\left(\frac{\operatorname{tr}(\rho^{2\alpha-1})}{s} + \frac{d^{2-2\alpha}}{d^2}\left(\frac{d^2}{s}\right)^{2\alpha}\right)\\
&\le O_\alpha\left(F^2\left[\frac{d^{1-1/\alpha}}{s} + \left(\frac{d^2}{s}\right)^{2\alpha}\right]\right),
\end{aligned} \tag{39}$$

where the last step uses Lemma 3.12 and $d^{2-2\alpha} \le F^2$.

For the second term, Equation (37) gives

$$\phi(T) = F + \sum_{j=1}^k c_j(\rho)T^j + r(T), \qquad |c_j(\rho)| \le O_\alpha(F), \qquad |r(T)| \le O_\alpha(FT^\alpha).$$

Because k is fixed, Lemma 3.11 and $\mathbf{Var}[r(T)] \le \mathbb{E}[r(T)^2]$ imply

$$\begin{aligned}
\mathop{\mathbf{Var}}_T[\phi(T)] &= \mathop{\mathbf{Var}}_T[\sum_{j=1}^k c_j(\rho)T^j + r(T)] \\
&\le O_\alpha\left(F^2\left[\sum_{j=1}^k \frac{1}{d^2}\left(\frac{d^2}{s}\right)^{2j} + \mathbb{E}[T^{2\alpha}]\right]\right) \\
&\le O_\alpha\left(F^2\left[\frac{d^2}{s^2} + \left(\frac{d^2}{s}\right)^{2\alpha}\right]\right), \qquad (40)
\end{aligned}$$

where Equation (40) uses Lemma 3.6, $d^2/s \le 1$ and k is a constant. Since $s \ge d^2$, $d^2/s^2 \le 1/s \le d^{1-1/\alpha}/s$. Combining Equation (38), Equation (39), Equation (40) proves Equation (28) and completes the proof of Lemma 3.4.

3.4 Technical lemmas

3.4.1 Hayashi's POVM

Lemma 3.6. *Let $|v\rangle$ be the outcome of the Hayashi's POVM applied on $|\psi\rangle^{\otimes s}$. Choose the phase of the outcome so that $\langle v|\psi\rangle \ge 0$. Then $|v\rangle = \sqrt{1-T}\,|\psi\rangle + \sqrt{T}\,|G\rangle$, where G is uniform on the unit sphere of $\psi^\perp$, T and G are independent, and $T \sim \mathrm{Beta}(d^2-1, s+1)$. For every fixed $q > 0$ and $s \ge d^2$,*

$$\mathbb{E}[T^q] = \frac{\Gamma(d^2-1+q)\Gamma(d^2+s)}{\Gamma(d^2-1)\Gamma(d^2+s+q)} \le O_q\left(\left(\frac{d^2}{s}\right)^q\right). \qquad (41)$$

For integer $j \ge 0$, $\mathbb{E}[T^j] = \mu_j(s)$, with μ_j as in Equation (25).

Proof. For a Haar random $|v\rangle$, one has $|\langle v|\psi\rangle|^2 \sim \mathrm{Beta}(1, d^2-1)$, with density function $f(x) = (d^2-1)(1-x)^{d^2-2}$. These facts follow, for example, by writing a Haar-random vector as a normalized standard complex Gaussian vector.

Now, let $|v\rangle$ be the outcome of Hayashi's POVM. Then $|v\rangle$'s density is obtained from Haar measure by multiplying by $\binom{d^2+s-1}{s}|\langle v|\psi\rangle|^{2s}$. Therefore, the $|\langle v|\psi\rangle|^2$ has density

$$\binom{d^2+s-1}{s}(d^2-1)x^s(1-x)^{d^2-2} = \frac{x^s(1-x)^{d^2-2}}{B(s+1, d^2-1)},$$

where $B(\cdot,\cdot)$ denotes the Euler beta function. Therefore $|\langle v|\psi\rangle|^2 \sim \mathrm{Beta}(s+1, d^2-1)$. This means $T = 1 - |\langle v|\psi\rangle|^2 \sim \mathrm{Beta}(d^2-1, s+1)$, and $|v\rangle = \sqrt{1-T}\,|\psi\rangle + \sqrt{T}\,|G\rangle$, where G is on the unit sphere of $\psi^\perp$. Furthermore, one can easily check that the law of G conditioned on any value of T is unitary invariant on $\psi^\perp$, which means G is uniformly distributed and independent of T.

The beta-integral identity [OLBC10, Eq. (5.12.1)] gives

$$\begin{aligned}
\mathbb{E}[T^q] &= \frac{1}{B(d^2-1, s+1)}\int_0^1 t^{d^2-2+q}(1-t)^s\,\mathrm{d}t \\
&= \frac{B(d^2-1+q, s+1)}{B(d^2-1, s+1)} \\
&= \frac{\Gamma(d^2-1+q)\Gamma(d^2+s)}{\Gamma(d^2-1)\Gamma(d^2+s+q)}.
\end{aligned}$$

To prove the required bound, first let $r \geq 1$ be an integer. Then

$$\mathbb{E}[T^r] = \prod_{k=0}^{r-1} \frac{d^2 - 1 + k}{d^2 + s + k} \leq \left(\frac{d^2 + r}{s}\right)^r \leq (r+1)^r \left(\frac{d^2}{s}\right)^r = O_r\left(\left(\frac{d^2}{s}\right)^r\right).$$

For general $q > 0$, write $q = m + \theta$, where $m = \lfloor q \rfloor$ and $0 \leq \theta < 1$. By Hölder's inequality,

$$\mathbb{E}[T^q] \leq (\mathbb{E}[T^m])^{1-\theta} \big(\mathbb{E}\big[T^{m+1}\big]\big)^{\theta} \leq O_q\left(\left(\frac{d^2}{s}\right)^q\right).$$

This proves Equation (41).

Finally, for an integer $j \geq 1$,

$$\mathbb{E}\big[T^j\big] = \prod_{r=0}^{j-1} \frac{d^2 - 1 + r}{d^2 + s + r} = \mu_j(s),$$

and the case $j = 0$ is immediate. □

3.4.2 Coefficients

Lemma 3.7. *The system Equation (26) has a unique solution, and*

$$\sum_{\ell=0}^{k} |a_\ell| + \sum_{\ell=0}^{k} a_\ell^2 \leq O_\alpha(1). \tag{42}$$

Proof. The factors

$$\prod_{r=0}^{j-1} (d^2 - 1 + r)$$

in Equation (25) depend only on j and are nonzero. It therefore suffices to consider the matrix

$$B_{j\ell} := \left[\prod_{r=0}^{j-1} (x_\ell + r)\right]^{-1}, \qquad x_\ell := d^2 + s_\ell,$$

and Equation (26) is thus equivalent to $Ba = (1, 0, \ldots, 0)^\mathsf{T}$.

Now, we prove B is non-singular. Multiply column ℓ of B by $\prod_{r=0}^{k-1}(x_\ell + r)$, then the entry in row j and column ℓ then becomes

$$\prod_{r=j}^{k-1} (x_\ell + r),$$

which is a monic polynomial in x_ℓ of degree $k - j$. After reversing the row order, the rows are evaluations of monic polynomials of degrees $0, 1, \ldots, k$. By elementary row operations (subtracting suitable linear combinations of the preceding rows), this matrix is transformed into the ordinary Vandermonde matrix $\{x_\ell^j\}_{0 \leq j, \ell \leq k}$. Thus its determinant is, up to the sign coming from reversing the rows, $\prod_{0 \leq \ell < h \leq k}(x_h - x_\ell)$. Since $x_0 < \cdots < x_k$, this determinant is nonzero, and hence the system has a unique solution.

For the uniform estimate, rescale row j of B by m^j and set $q := \frac{d^2}{m} \in (0, 1]$, $\tau := \frac{1}{m} \in (0, 1/4]$. The rescaled matrix C has entries

$$C_{j\ell} = \prod_{r=0}^{j-1} \frac{1}{2^\ell + q + r\tau}.$$

The same Vandermonde calculation gives

$$|\det C| = \frac{\prod_{0\le \ell<h\le k}(2^h-2^\ell)}{\prod_{\ell=0}^{k}\prod_{r=0}^{k-1}(2^\ell+q+r\tau)}.$$

For $(q,\tau)\in[0,1]\times[0,1/4]$, the absolute values of entries of C are uniformly upper bounded, while the displayed determinant is bounded below by a positive constant depending only on k. Hence the absolute values of entries of C^{-1} are uniformly upper bounded. Note that a is also the solution of $Ca=(1,0,\ldots,0)^\mathsf{T}$ since the row rescaling R does not change the right hand side, i.e., let R denote the row rescaling matrix, then

$$Ca = RBa = R(1,0,\ldots,0)^\mathsf{T} = (1,0,\ldots,0)^\mathsf{T},$$

each a_ℓ is bounded by a constant depending only on k, and therefore only on α. As the number of coefficients is fixed, Equation (42) follows. □

3.4.3 Schatten norm expansion and moments

Lemma 3.8 ([PS14])**.** *Fix $p>2$. There are constants C_p and $C_{p,j}$ such that, for every $d\ge 1$ and every $M\in\mathbb{C}^{d\times d}$ with $\|M\|_p\le 1$, there are symmetric real multilinear forms $L_{j,M}:(\mathbb{C}^{d\times d})^j\to\mathbb{R}$ for $1\le j\le\lceil p\rceil-1$, where $\mathbb{C}^{d\times d}$ is viewed as a real vector space, satisfying*

$$|L_{j,M}(H_1,\ldots,H_j)| \le C_{p,j}\|M\|_p^{p-j}\prod_{i=1}^{j}\|H_i\|_p \tag{43}$$

for any $H_1,\ldots,H_j\in\mathbb{C}^{d\times d}$, and for every $H\in\mathbb{C}^{d\times d}$ with $\|H\|_p\le 1$, it holds that

$$\|M+H\|_p^p = \|M\|_p^p + \sum_{j=1}^{\lceil p\rceil-1} L_{j,M}(H,\ldots,H) + \mathcal{R}_p(M,H), \qquad |\mathcal{R}_p(M,H)|\le C_p\|H\|_p^p. \tag{44}$$

Proof. Let $r=\lceil p\rceil-1$. Since $r<p\le r+1$, [PS14, Theorem 16] provides symmetric real multilinear forms $\delta_A^{(j)}$ for $1\le j\le r$, which, by [PS14, Equations (23)–(24) and Theorem 18], satisfy

$$|\delta_A^{(j)}(V_1,\ldots,V_j)| \le C_{p,j}\|A\|_p^{p-j}\prod_{i=1}^{j}\|V_i\|_p.$$

Although [PS14, Theorem 16] states its remainder only as $\|V\|_p\to 0$, the final step of its proof gives

$$\left|\|A_1\|_p^p - \|A_0\|_p^p - \sum_{j=1}^{r}\delta_{A_0}^{(j)}(V,\ldots,V)\right| \le C_p\|V\|_p^p, \qquad V=A_1-A_0,$$

whenever $\|A_0\|_p,\|A_1\|_p\le 1$; see [PS14, Equation (29), Theorem 14, Remark 15, and pp. 471–473].

For the matrices in the lemma, take $A_0=M/2$, $V=H/2$, and $A_1=(M+H)/2$, which satisfy $\|A_0\|_p\le 1$ and $\|A_1\|_p\le 1$. The proof is completed by setting

$$L_{j,M}(H_1,\ldots,H_j) := 2^{p-j}\delta_{M/2}^{(j)}(H_1,\ldots,H_j).$$

□

Lemma 3.9. *Let $M \in \mathbb{C}^{d\times d}$ satisfy $\|M\|_{\mathrm{F}} = 1$, and let G be uniform on the unit sphere of the complex Frobenius hyperplane $M^{\perp} = \{X \in \mathbb{C}^{d\times d} \mid \langle M, X\rangle_{\mathrm{F}} = 0\}$. For every fixed $q \geq 2$ and $u > 0$,*

$$\left(\mathbb{E}\left[\|G\|_q^u\right]\right)^{1/u} \leq O_{q,u}(d^{1/q-1/2}). \tag{45}$$

In particular,

$$\mathbb{E}\left[\|G\|_{2\alpha}^{2\alpha}\right] \leq O_\alpha\left(d^{1-\alpha}\right) \leq O_\alpha(\mathrm{F}_\alpha(\rho)) \tag{46}$$

whenever $\rho = MM^{\dagger}$.

Proof. Let $W \in \mathbb{C}^{d\times d}$ have independent standard complex Gaussian entries, i.e., $W_{ij} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and set $Z := W - \langle M, W\rangle_{\mathrm{F}} M$. Then Z is a standard complex Gaussian vector in $M^{\perp}$. Thus, in any complex orthonormal basis $E_1, \ldots, E_{d^2-1}$ of $M^{\perp}$,

$$Z = \sum_{j=1}^{d^2-1} \zeta_j E_j, \qquad \zeta_1, \ldots, \zeta_{d^2-1} \sim \mathcal{N}_{\mathbb{C}}(0, 1).$$

Set $\widetilde{G} := Z/\|Z\|_{\mathrm{F}}$. As noted in [AS17, proof of Proposition 6.34, p. 171], $\widetilde{G}$ is uniform on the unit sphere of $M^{\perp}$ and is independent of $\|Z\|_{\mathrm{F}}$; in particular, $\widetilde{G}$ has the same distribution as G. Therefore,

$$\mathbb{E}\left[\|Z\|_q^u\right] = \mathbb{E}\left[\|Z\|_{\mathrm{F}}^u \|\widetilde{G}\|_q^u\right] = \mathbb{E}[\|Z\|_{\mathrm{F}}^u]\, \mathbb{E}\left[\|G\|_q^u\right],$$

and hence

$$(\mathbb{E}\left[\|G\|_q^u\right])^{1/u} = \frac{(\mathbb{E}\left[\|Z\|_q^u\right])^{1/u}}{(\mathbb{E}\left[\|Z\|_{\mathrm{F}}^u\right])^{1/u}}. \tag{47}$$

Note that $|\zeta_j|^2 \sim \mathrm{Exp}(1)$ (cf. [AS17, Equation (A.2), p. 307]), and thus

$$\|Z\|_{\mathrm{F}}^2 = \sum_{j=1}^{d^2-1} |\zeta_j|^2 \sim \mathrm{Gamma}(d^2 - 1, 1).$$

Now that $q \geq 2$ and $\|M\|_q \leq \|M\|_{\mathrm{F}} = 1$, we have

$$\|Z\|_q \leq \|W\|_q + |\langle M, W\rangle_{\mathrm{F}}|\|M\|_q \leq d^{1/q}\|W\|_{\mathrm{op}} + |\langle M, W\rangle_{\mathrm{F}}|. \tag{48}$$

For $u > 0$, put $K_u := 2^{\max\{u-1,0\}}$, so that $(x + y)^u \leq K_u(x^u + y^u)$ for all $x, y \geq 0$. By [AS17, Proposition 6.33, Equation (6.42), p. 169], we have

$$\Pr\left[\|W\|_{\mathrm{op}} > 2\sqrt{d} + t\right] \leq \frac{1}{2}e^{-t^2}, \qquad t > 0.$$

Let $Y := \max\{\|W\|_{\mathrm{op}} - 2\sqrt{d}, 0\}$. Then,

$$\begin{aligned}\mathbb{E}[Y^u] &= \int_0^\infty ut^{u-1} \Pr[Y > t]\, \mathrm{d}t \\ &\leq \frac{u}{2}\int_0^\infty t^{u-1}e^{-t^2}\, \mathrm{d}t = \frac{1}{2}\Gamma\left(1 + \frac{u}{2}\right).\end{aligned}$$

Since $\|W\|_{\mathrm{op}} \leq 2\sqrt{d} + Y$, it follows that

$$\mathbb{E}\left[\|W\|_{\mathrm{op}}^u\right] \leq K_u \left(2^u d^{u/2} + \frac{1}{2}\Gamma\left(1 + \frac{u}{2}\right)\right) \leq O_u\left(d^{u/2}\right), \tag{49}$$

and hence $(\mathbb{E}[\|W\|_{\mathrm{op}}^u])^{1/u} \le O_u(\sqrt{d})$.

Next, note that $\langle M, W\rangle_{\mathrm{F}} \sim \mathcal{N}_{\mathbb{C}}(0,1)$. Therefore, [AS17, Equation (A.2), p. 307] gives

$$\mathbb{E}[|\langle M, W\rangle_{\mathrm{F}}|^u] = \Gamma\Big(1+\frac{u}{2}\Big).$$

By Equations (48) and (49), we have

$$\mathbb{E}\big[\|Z\|_q^u\big] \le K_u\left(d^{u/q}\,\mathbb{E}\big[\|W\|_{\mathrm{op}}^u\big] + \mathbb{E}[|\langle M, W\rangle_{\mathrm{F}}|^u]\right) \le O_{q,u}\Big(d^{u(1/q+1/2)}\Big),$$

and hence $(\mathbb{E}[\|Z\|_q^u])^{1/u} \le O_{q,u}(d^{1/q+1/2})$. Since $M^\perp$ has complex dimension d^2-1, the gamma-moment formula gives

$$(\mathbb{E}[\|Z\|_{\mathrm{F}}^u])^{1/u} = \left(\frac{\Gamma(d^2-1+u/2)}{\Gamma(d^2-1)}\right)^{1/u} = \Theta_u(d).$$

Substituting these two estimates into Equation (47) yields

$$\Big(\mathbb{E}\big[\|G\|_q^u\big]\Big)^{1/u} \le O_{q,u}\Big(d^{1/q-1/2}\Big).$$

□

3.4.4 Some facts

Lemma 3.10. *For every $\beta > 0$, there is a polynomial J_β of degree at most $\lceil\beta\rceil - 1$ such that*

$$|(1-t)^\beta - J_\beta(t)| \le O_\beta(t^\beta), \qquad (0 \le t \le 1).$$

One may take the Taylor polynomial at $t = 0$ through degree $\lceil\beta\rceil - 1$.

Proof. For $0 \le t \le 1/2$, Taylor's theorem gives a remainder of order $t^{\lceil\beta\rceil}$, which is at most t^β. For $1/2 \le t \le 1$, both the function and the fixed polynomial are bounded, while $t^\beta \ge 2^{-\beta}$. □

Lemma 3.11. *Let $T \sim \mathrm{Beta}(d^2-1, s+1)$, where $d \ge 2$ and $s \ge d^2$. For every fixed integer $j \ge 1$,*

$$\mathbf{Var}[T^j] \le O_j\left(\frac{1}{d^2}\left(\frac{d^2}{s}\right)^{2j}\right). \tag{50}$$

Proof. Set $a = d^2-1$, $b = s+1$, and $M_j := \mathbb{E}[T^j]$. The beta moment formula gives

$$\frac{M_{2j}}{M_j^2} = \prod_{r=0}^{j-1}\frac{a+j+r}{a+r}\,\frac{a+b+r}{a+b+j+r} \le \prod_{r=0}^{j-1}\left(1+\frac{j}{a+r}\right) \le 1 + O_j\left(\frac{1}{d^2}\right).$$

Thus $\mathbf{Var}[T^j] \le O_j(M_j^2/d^2)$. Also $M_j \le O_j((d^2/s)^j)$ by Equation (41), proving the claim. □

Lemma 3.12. *For every state ρ and every $\alpha > 1$,*

$$\mathrm{tr}(\rho^{2\alpha-1}) \le \mathrm{F}_\alpha(\rho)^{2-1/\alpha} \le \mathrm{F}_\alpha(\rho)^2 d^{1-1/\alpha}. \tag{51}$$

Proof. Let $\lambda_{\max}$ be the largest eigenvalue of ρ. Then

$$\mathrm{tr}(\rho^{2\alpha-1}) \le \lambda_{\max}^{\alpha-1}\mathrm{F}_\alpha(\rho) \le \mathrm{F}_\alpha(\rho)^{(\alpha-1)/\alpha}\mathrm{F}_\alpha(\rho),$$

because $\lambda_{\max}^\alpha \le \mathrm{F}_\alpha(\rho)$. The second inequality in Equation (51) follows from $\mathrm{F}_\alpha(\rho) \ge d^{1-\alpha}$. □

Proof. The assumption gives $\widehat{F}/F \geq e^{-\gamma}$ and $\widehat{F}/F \leq 2 - e^{-\gamma}$. Since $e^{\gamma} + e^{-\gamma} \geq 2$, one has $2 - e^{-\gamma} \leq e^{\gamma}$. Thus $e^{-\gamma} \leq \widehat{F}/F \leq e^{\gamma}$, and taking logarithms proves the claim. $\square$

Lemma 3.13. *Let H be a real inner-product space of dimension $m \geq 2$, and let*

$$\mathbb{S}(H) := \{x \in H : \|x\| = 1\}.$$

If G is uniformly distributed on $\mathbb{S}(H)$, then every continuously differentiable function $f : \mathbb{S}(H) \to \mathbb{R}$ satisfies

$$\mathbf{Var}[f(G)] \leq \frac{1}{m-1}\,\mathbb{E}\Big[\big\|\nabla_{\mathbb{S}(H)} f(G)\big\|^2\Big]. \tag{52}$$

Here $\nabla_{\mathbb{S}(H)} f$ denotes the tangential gradient of f along the sphere.

In particular, if $\widetilde{f}$ is a continuously differentiable extension of f to a neighborhood of $\mathbb{S}(H)$ in H, then

$$\mathbf{Var}[f(G)] \leq \frac{1}{m-1}\,\mathbb{E}\left[\left\|\nabla \widetilde{f}(G)\right\|^2\right], \tag{53}$$

where ∇ is simply the Euclidean gradient with respect to the inner product on H.

Proof. After identifying H isometrically with $\mathbb{R}^m$, Equation (52) is the classical Poincaré inequality on $\mathbb{S}^{m-1}$, applied to $f - \mathbb{E}[f(G)]$; see [BCPR23, Proposition 2.3].

For the second assertion, observe that $\nabla_{\mathbb{S}(H)} f(x)$ is the orthogonal projection of $\nabla_H \widetilde{f}(x)$ onto the tangent space $T_x\mathbb{S}(H)$. Hence

$$\big\|\nabla_{\mathbb{S}(H)} f(x)\big\| \leq \left\|\nabla \widetilde{f}(x)\right\|,$$

and Equation (53) follows from Equation (52). $\square$

Acknowledgment

The authors used Large Language Models as AI-assisted research and writing tools throughout the preparation of this paper. The tools were used to help brainstorm ideas and explore proof strategies. Portions of the paper text were redrafted or modified with AI assistance across all sections. All final mathematical claims, algorithms, proofs, citations, and wording were reviewed, edited, and validated by the authors. The authors assume responsibility for all content of the paper.

References

[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, 4, 5

[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. 3

[AS17] Guillaume Aubrun and Stanisław J. Szarek. *Alice and Bob Meet Banach: The Interface of Asymptotic Geometric Analysis and Quantum Information Theory*, volume 223 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2017. doi:10.1090/surv/223. 26, 27

[BCPR23] Chiara Bianchini, Andrea Colesanti, Daniele Pagnini, and Alberto Roncoroni. On p-brunn–minkowski inequalities for intrinsic volumes, with $0 \leq p < 1$. *Mathematische Annalen*, 387:321–352, 2023. doi:10.1007/s00208-022-02454-0. 28

[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. 4, 5

[BDKR05] Tugkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005. doi:10.1137/S0097539702403645. 3

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997. doi:10.1007/978-1-4612-0653-8. 11

[BMW16] Mohammad Bavarian, Saeed Mehraban, and John Wright. Learning entropy. A manuscript on von Neumann entropy estimation, private communication, 2016. 3, 5

[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

[CLW26] Kean Chen, Yupan Liu, and Qisheng Wang. Trace estimation of quantum state powers: Sample complexity and computational hardness. ArXiv preprints, 2026. arXiv:2505.09563v2. 3, 4

[CW25] Kean Chen and Qisheng Wang. Improved sample upper and lower bounds for trace estimation of quantum state powers. In *The 38th Annual Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 1008–1028, 2025. URL: https://proceedings.mlr.press/v291/chen25d.html. 3, 4, 5

[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. 4, 5

[CWYZ26] Kean Chen, Qisheng Wang, Zhan Yu, and Zhicheng Zhang. Simultaneous estimation of nonlinear functionals of a quantum state. *IEEE Transactions on Information Theory*, 2026. doi:10.1109/TIT.2026.3699531. 4, 5

[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. 4, 5

[FOW26] Marco Fanizza, Ryan O'Donnell, and Chirag Wadhwa. Spectrum estimation is almost as hard as tomography. ArXiv preprints, 2026. arXiv:2607.29680. 3

[GHS21] Tom Gur, Min-Hsiu Hsieh, and Sathyawageeswar Subramanian. Sublinear quantum algorithms for estimating von Neumann entropy. ArXiv e-prints, 2021. arXiv:2111.11139. 6

[GHYZ26] Weiyuan Gong, Jonas Haferkamp, Qi Ye, and Zhihan Zhang. On the sample complexity of purity and inner product estimation. *Physical Review Research*, 2026. doi:10.1103/kh2j-jkn6. 4, 5

[GL20] András Gilyén and Tongyang Li. Distributional property testing in a quantum world. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, pages 25:1–25:19, 2020. doi:10.4230/LIPIcs.ITCS.2020.25. 6

[GW26] Minbo Gao and Qisheng Wang. Breaking the quadratic barrier for von Neumann entropy estimation, 2026. arXiv:2608.11151. 3

[Hay98] Masahito Hayashi. Asymptotic estimation theory for a finite-dimensional pure state model. *Journal of Physics A: Mathematical and General*, 31(20):4633–4655, 1998. doi:10.1088/0305-4470/31/20/006. 6, 18

[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. 3

[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. 3

[LS01] Adrian S. Lewis and Hristo S. Sendov. Twice differentiable spectral functions. *SIAM Journal on Matrix Analysis and Applications*, 23(2):368–386, 2001. doi:10.1137/S089547980036838X. 11

[LW26] Yupan Liu and Qisheng Wang. On estimating the trace of quantum state powers. *IEEE Transactions on Information Theory*, 72(8):5830–5862, 2026. doi:10.1109/TIT.2026.3683891. 4, 6

[OLBC10] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, 2010. See Section 5.12(i), Equation (5.12.1). 23

[OW17] Ryan O'Donnell and John Wright. Efficient quantum tomography II. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 962–974, 2017. doi:10.1145/3055399.3055454. 3, 5

[Pan03] Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003. doi:10.1162/089976603321780272. 3

[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. 3

[PS14] Denis Potapov and Fedor Sukochev. Fréchet differentiability of $\mathcal{S}^p$ norms. *Advances in Mathematics*, 262:436–475, 2014. doi:10.1016/j.aim.2014.05.011. 25

[PSTW26] Angelos Pelecanos, Jack Spilecki, Ewin Tang, and John Wright. The Keyl-Werner algorithm is not optimal for spectrum estimation, 2026. arXiv:2607.27117. 5, 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. 6

[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

[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. 6

[Sha48a] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x. 3

[Sha48b] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656, 1948. doi:10.1002/j.1538-7305.1948.tb00917.x. 3

[TKR+10] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete. The χ^2-divergence and mixing times of quantum Markov processes. *Journal of Mathematical Physics*, 51(12):122201, 2010. doi:10.1063/1.3511335. 12

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

[TWZ25] Ewin Tang, John Wright, and Mark Zhandry. Conjugate queries can help. *arXiv preprint arXiv:2510.07622*, 2025. 6, 17

[Val11] Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011. doi:10.1137/080734066. 3

[vN27] J. v. Neumann. Thermodynamik quantummechanischer gesamheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1927:273–291, 1927. URL: http://eudml.org/doc/59231. 3

[vN32] John von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer, 1932. URL: http://eudml.org/doc/203794. 3

[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. 3

[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. 3

[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. 3

[Wan25] Qisheng Wang. Information-theoretic lower bounds for approximating monomials via optimal quantum Tsallis entropy estimation. ArXiv preprint, 2025. arXiv:2509.03496. 6

[Wan26] Qisheng Wang. A lower bound framework for quantum functional estimation, 2026. arXiv:2608.02600. 3, 4, 5

[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. 6

[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. 3

[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. 3

[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. 6