Proceedings of Machine Learning Research vol 336:1–71, 2026 39th Annual Conference on Learning Theory

Instance-optimal high-precision shadow tomography with few-copy measurements: A metrological approach

Senrui Chen CSENRUI@GMAIL.COM
California Institute of Technology
Weiyuan Gong WGONG@G.HARVARD.EDU
Harvard University
Sisi Zhou SISI.ZHOU26@GMAIL.COM
Perimeter Institute and University of Waterloo

Editors: Steve Hanneke and Tor Lattimore

Abstract

We give the first instance-optimal sample complexity bounds for shadow tomography using few-copy measurements in the high-precision regime. More concretely, we study the problem of learning expectation values of a given set of observables of an unknown quantum state to precision ε in L_p-norm, using (possibly adaptive) measurements that act on one or a few copies at a time, and we are interested in the regime that ε is below some concrete and potentially dimension-dependent threshold. In this setup, we prove the necessary and sufficient number of copies, for any given set of observables, is characterized by a simple optimization formula involving a quadratic form of the inverse Fisher information matrix up to a logarithmic factor. Our results establish a rigorous correspondence between quantum learning and quantum metrology.

Keywords: Shadow tomography, High precision, Metrology, Fisher information

1. Introduction

A fundamental task in quantum information is to characterize an unknown or partly unknown state, with applications in *e.g.* quantum sensing (Degen et al., 2017; Pirandola et al., 2018), quantum algorithms (Dalzell et al., 2025), as well as benchmarking noisy quantum devices (Harper et al., 2020; Hashim et al., 2024). There are two fields of research that have been devoted to conquering this task: One is Quantum Learning; the other is Quantum Metrology. Despite the same over-arching goal of understanding how efficiently one can extract information about an unknown quantum system, the languages, techniques, and communities are surprisingly different.

Quantum metrology, or quantum estimation theory, is a topic with a long history and is widely applied in experiments (Giovannetti et al., 2011; Degen et al., 2017; Pezze et al., 2018; Pirandola et al., 2018). As a typical setting, a quantum state is parameterized by one or many unknown parameters. The goal is to understand how precisely the parameters can be determined as the number of available state copies goes to infinity. Modern theoretical quantum metrology research mostly relies on (quantum) Fisher information and the (quantum) Cramér–Rao bound, which are powerful tools in asymptotic statistics (Van der Vaart, 2000).

Quantum learning is a younger field, strongly influenced by the computer science and machine learning community (Arunachalam and De Wolf, 2017). One typical question, known as shadow tomography (Aaronson, 2018), asks how many copies of an d-dimensional unknown state are needed to estimate a set of observables to certain precision ε. Crucially, instead of investigating the $\varepsilon \to 0$ limit, one is interested in the scaling of complexity with finite ε and d. This falls into the regime of non-asymptotic statistics, and people are using very different methods than those used in quantum metrology research.

Nevertheless, many have wondered about the following question:

Can we establish a rigorous correspondence between quantum metrology and quantum learning?

We provide an affirmative answer by using quantum metrological approaches to solve an important open problem in quantum learning theory: namely, the instance-optimal sample complexity for high-precision shadow tomography with few-copy measurements. At a high level, the problem is to estimate many given observables of an unknown quantum states to ε precision in L_p norm, with the restriction that each round of measurement acts upon one or a few number of copies. Here, "instance-optimal" means the bounds depend on the set of observables, and "high-precision" means we restrict ε to be smaller than a concrete threshold that may depend on the dimension of the Hilbert space and the specific set of observables. For this task, we essentially show that an intuitive bound one would expect using Fisher information gives a tight characterization of the sample complexity. Our results show that quantum metrology and learning are not only closely related in concept, but has exact mathematical correspondence in an appropriate regime.

2. Results

2.1. Six problems: learning, estimation and distinguishing

Consider a d-dimensional Hilbert space. Let $\{O_i\}_{i=1}^m$ be a set of linearly-independent (thus $m \leq d^2-1$) traceless Hermitian operators that we want to learn. Define the dual operator basis $\{Q_a\}_{a\in A}\cup \{T_b\}_{b\in B}$ which form a complete basis of traceless Hermitian observables satisfying

$$\operatorname{tr}(O_iQ_a) = d\delta_{ia}, \quad \forall i, a \in A, \quad \operatorname{tr}(O_iT_b) = 0, \quad \forall i \in A, b \in B. \tag{2.1}$$

We use $A = \{1, \ldots, m\}$ and $B = \{m+1, \ldots, d^2-1\}$ to denote the set of indices for Q_a and T_b. Now, fix any (known) reference quantum state ρ_0, one can always parameterize any quantum state ρ by

$$\rho_{\theta,\varphi} := \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b. \tag{2.2}$$

With these definitions, we are ready to introduce the following problems about quantum state learning, estimation, and distinguishing. $p \in [1, \infty]$ is a variable index in the following problems.

Problem 1 (Learning of observables with p-norm error) *Given N i.i.d. copies of a quantum state ρ, find estimators $\{\hat{o}_i\}_{i=1}^m$ such that $(\sum_{i=1}^m |\operatorname{tr}(O_i\rho) - \hat{o}_i|^p)^{1/p} < \varepsilon$ with high probability.*

Problem 2 (Estimation of parameters with p-norm error) *Given N i.i.d. copies of a parameterized quantum state $\rho_{\theta,\varphi}$ for some known (but arbitrary) choice of ρ_0, $\{Q_a\}_{a\in A}$ and $\{T_b\}_{b\in B}$, find estimators $\{\hat{\theta}_i\}_{i=1}^m$ such that $(\sum_{i=1}^m |\hat{\theta}_i - \theta_i|^p)^{1/p} < \varepsilon$ with high probability.*

Problem 3 (Distinguishing between one and many states) *Given some known (but arbitrary) choice of ρ_0, $\{Q_a\}_{a\in A}$ and $\{T_b\}_{b\in B}$, and N identical copies of a quantum state that is either ρ_0 or $\rho_{\theta,\varphi}$ with equal probability, where $\theta \in \mathbb{R}^m$ and $\varphi \in \mathbb{R}^{d^2-m-1}$ can be arbitrary unknown vectors such that $\rho_{\theta,\varphi}$ is well-defined and $\|\theta\|_p = 3\varepsilon$, distinguish the two cases with high probability.*

$$\begin{tikzcd}
\begin{array}{c}\text{Learning}\\\text{(Problem 1)}\end{array} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \begin{array}{c}\text{Estimation}\\\text{(Problem 2)}\end{array} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \begin{array}{c}\text{Distinguishing}\\\text{(Problem 3)}\end{array} \arrow[d, Rightarrow] \\
\begin{array}{c}\text{Oblivious Learning}\\\text{(Problem 1')}\end{array} \arrow[r, Leftrightarrow] & \begin{array}{c}\text{Oblivious Estimation}\\\text{(Problem 2')}\end{array} \arrow[r, Rightarrow] & \begin{array}{c}\text{Oblivious Distinguishing}\\\text{(Problem 3')}\end{array}
\end{tikzcd}$$

Figure 1: Relationship between six problems. $\Leftrightarrow$ means two problems are equivalent, and $\Leftarrow$ means one problem is no harder than (reduces to) the other. Note that within each problem, increasing p (or decreasing q) will not increase the hardness of the problem.

Problem 1 is a standard quantum state learning scenario, where $p = \infty$ is traditionally known as shadow tomography where expectation values of observables are learned up to a constant additive error. In our work, we extend the discussion to a more general form of shadow tomography where error is measured in p-norm norm for any $p \in [1, \infty]$, and obtain tight bounds on the sample complexity in the high-precision regime where ε is sufficiently small. We assume the choices of observables $\{O_i\}_{i=1}^m$ are known to the experimentalists prior to the experiments, which allows optimization of the learning algorithms based on the knowledge of target observables. As a result, the bounds we obtain are *instance-optimal*, as functions of observables $\{O_i\}_{i=1}^m$, which capture different levels of difficulties when learning different types of observables.

To put the quantum state learning problem in the context of quantum state estimation where techniques from quantum metrology are available, we first observe that Problem 1 is equivalent to the state estimation problem Problem 2.

Proof [Equivalence between Problem 1 and Problem 2] We note that any state ρ can be parameterized as Eq. (2.2) because ρ_0 and $\{Q_a\}_{a\in A} \cup \{T_b\}_{b\in B}$ form a complete basis of d-dimensional Hermitian operators. Furthermore, $\theta_a = \mathrm{tr}(\rho O_a) - \mathrm{tr}(\rho_0 O_a)$. Thus, estimating θ_a to p-norm error is equivalent to estimating $\mathrm{tr}(\rho O_a)$ up to p-norm error because they differ only by a known, constant vector. ■

Hypothesis testing or distinguishing tasks are in general easier than learning (or estimation) tasks, and they conveniently provide lower bounds on the resource required for learning tasks from an information-theoretic point of view. The intuition is successful learning of quantum state properties can be used to distinguish different types of states. In our case, we consider a many-versus-one distinguishing task where a reference state ρ_0 is to be distinguished from $\rho_{\theta,\varphi}$ which is close to ρ_0 under our p-norm metric.

Proof [Problem 3 is no harder than Problem 2] Consider the unknown state ρ given in Problem 3. Let $\hat{\theta}$ be the estimator constructed from Problem 2. If $\rho = \rho_0$, $\|\hat{\theta}\|_p < \varepsilon$ with high probability; and if $\rho = \rho_{\theta,\varphi}$, $\|\hat{\theta} - \theta\|_p < \varepsilon$ and $\|\hat{\theta}\|_p \geq \|\theta\|_p - \|\hat{\theta} - \theta\|_p > 2\varepsilon$ with high probability. Thus, by deciding whether $\|\hat{\theta}\|_p$ is closer to 0 or 3ε, Problem 3 can be solved with high probability by $\hat{\theta}$. ■

As we will see later, when $p = \infty$ and ε is sufficiently small, Problem 3 is as hard as Problem 2 up to a logarithmic overhead. It implies the distinguishing capability can almost capture the learning/estimation capability in certain regimes which in general does not hold.

Next, we consider a different but also practically relevant setting which we call *oblivious* learning or oblivious estimation. Here, instead of aiming at precisely estimating all expectation values of observables, we need to only estimate a single observable chosen arbitrarily from a set which we call the L^q-ellipsoid of $\{O_i\}_{i=1}^m$,

$$\left\{ O_\alpha := \sum_{i=1}^m \alpha_i O_i, \quad \forall \alpha \in \mathbb{R}^m, \|\alpha\|_q = 1 \right\}, \tag{2.3}$$

where $q \in [1,\infty]$ satisfying $1/p + 1/q = 1$ is the dual index to p. The choice of O_α (i.e. the value of α) will only be revealed after all quantum measurements are performed, forbidding any informed optimization of the measurement prior to data processing—which explains the name "oblivious". Below we have the three oblivious versions of the previously defined learning, estimation, and distinguishing problems.

Problem 1' (Oblivious learning of an observable from L_q-ellipsoid) *Given N i.i.d. copies of quantum state ρ and an arbitrary $\alpha \in \mathbb{R}^m$ satisfying $\|\alpha\|_q \le 1$, where the values of α is revealed after all quantum measurements are performed, find an estimators $\hat{o}_\alpha$ such that $|\operatorname{tr}(\sum_{i=1}^m \alpha_i O_i \rho) - \hat{o}_\alpha| < \varepsilon$ with high probability.*

Problem 2' (Oblivious estimation of a linear function of parameters) *Given N i.i.d. copies of the parameterized quantum state $\rho_{\theta,\varphi}$ described by some known (but arbitrary) choice of ρ_0, $\{Q_a\}_{a\in A}$ and $\{T_b\}_{b\in B}$ and an arbitrary $\alpha \in \mathbb{R}^m$ satisfying $\|\alpha\|_q \le 1$, with the values of α revealed after all quantum measurements are performed, find an estimator $\hat{\theta}_\alpha$ such that $|\theta_\alpha - \hat{\theta}_\alpha| < \varepsilon$ with high probability where $\theta_\alpha := \alpha \cdot \theta$.*

Problem 3' (Oblivious distinguishing between two states) *Given some known (but arbitrary) choice of ρ_0, $\{Q_a\}_{a\in A}$ and $\{T_b\}_{b\in B}$ and N i.i.d. copies of a quantum state that is either ρ_0 or $\rho_{\theta,\varphi}$ with equal probability, where $\theta \in \mathbb{R}^m$ and $\varphi \in \mathbb{R}^{d^2-m-1}$ can be arbitrary unknown vectors such that $\rho_{\theta,\varphi}$ is well-defined and $\|\theta\|_p = 3\varepsilon$, with the values of (θ,φ) revealed after all quantum measurements are performed, distinguish the two cases with high probability.*

Analogous to the previous proof, we observe that Problem 1' is equivalent to Problem 2', and is no easier than Problem 3'.

Proof [Problem 3' is no harder than Problem 2'] Let (θ,φ) be the revealed parameters and ρ be the unknown state in Problem 3'. Pick $\alpha \in \mathbb{R}^m$ such that $\alpha \cdot \theta = 3\varepsilon$ and $\|\alpha\|_q = 1$. Let $\hat{\theta}_\alpha$ be the estimator constructed from Problem 2'. If $\rho = \rho_0$, $|\hat{\theta}_\alpha| < \varepsilon$ with high probability; and if $\rho = \rho_{\theta,\varphi}$, $|\alpha \cdot \theta - \hat{\theta}_\alpha| < \varepsilon$ and $|\hat{\theta}_\alpha| \ge |\alpha \cdot \theta| - |\hat{\theta}_\alpha - \alpha \cdot \theta| > 2\varepsilon$ with high probability. Thus, by deciding whether $|\hat{\theta}_\alpha|$ is closer to 0 or 3ε, Problem 3 can be solved with high probability by $\hat{\theta}_\alpha$. ■

Moreover, the oblivious versions of tasks described above are no harder than the original learning, estimation and distinguishing tasks. While Problem 3' is trivially no harder than Problem 3, the relationship between Problem 2' and Problem 2 can be seen from the following.

Proof [Problem 2' is no harder than Problem 2] Consider Problem 2 and let $\hat{\theta}$ be an estimator satisfying $\|\hat{\theta} - \theta\|_p < \varepsilon$ with high probability. Then $\hat{\theta}_\alpha = \alpha \cdot \hat{\theta}$ solves Problem 2' because $|\alpha \cdot \theta - \alpha \cdot \hat{\theta}| \le \|\alpha\|_q \|\hat{\theta} - \theta\|_p < \varepsilon$ using Hölder's inequality. ■

Oblivious estimation fits well into the "measure first, ask questions later" regime where the observable of interest is determined after the experiments. In our case, O_α can be chosen freely as any linear combination of $\{O_i\}_{i=1}^m$. As a result, the oblivious estimation task is also challenging to solve. As we will see later when $p = \infty$, the oblivious estimation and the original shadow estimation tasks have the same sample complexity up to logarithmic overhead.

2.2. Oblivious estimation from L^q-ellipsoid

In this section, we summarize the results we obtain on oblivious estimation and distinguishing tasks (see also Table 1). In particular, we will show tight upper and lower bounds on the sample complexity in the high-precision regime, i.e. when the target precision ε is sufficiently small.

Theorem 1 (Oblivious estimation with one-copy measurements, informal) *Given any $p \in [1, \infty]$, to solve Problem 1', Problem 2' or Problem 3' with (adaptive) single-copy measurement protocols, there exists a threshold on the target precision ε below which*

$$N = \Theta\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right) \tag{2.4}$$

copies of quantum states ρ are necessary and sufficient. Here $\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)$ is a positive function of the observables defined by

$$\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m) := \inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\substack{\|\alpha\|_q \le 1 \\ \alpha \in \mathbb{R}^m}} \alpha^\top (I(\rho_0, M)^{-1})_{AA} \alpha, \tag{2.5}$$

where $I(\rho_0, M)$ is the Fisher information matrix (FIM) $I(\rho_{\theta,\varphi}, M)$ of measuring $\rho_{\theta,\varphi}$ with measurement M at $(\theta, \varphi) = (0, 0)$, $\mathcal{M}$ is the set of all single-copy measurements, $\mathcal{S}^\circ$ is the set of all full-rank density matrices and AA denotes the upper-left matrix block of $I(\rho_0, M)$ when indices are restricted to $a \in A$.

Theorem 1 provides tight bounds on the sample complexity required to obliviously estimate observables from the L^q-ellipsoid of $\{O_i\}_{i=1}^m$. Here the function $\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)$ fully characterizes the sample complexity required to perform oblivious estimation or distinguishing tasks in the high-precision regime. To understand the operational meaning of $\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)$, we first note that the FIM

$$I(\rho_0, M)_{cc'} = \sum_x \frac{\mathrm{tr}(R_c M_x)\,\mathrm{tr}(R_{c'} M_x)}{\mathrm{tr}(\rho_0 M_x)}, \qquad R_c = \begin{cases} Q_a & c = a \in A, \\ T_b & c = b \in B. \end{cases} \tag{2.6}$$

characterizes the amount of information a parameterized quantum state contains about unknown parameters that can be extracted from quantum measurement $M = \{M_x\}_x$. Its inverse is considered an asymptotically attainable lower bound on the variance of estimators by the Cramér–Rao bound (Kay, 1993; Lehmann and Casella, 2006). Here the scalar $\alpha^\top (I(\rho_0, M)^{-1})_{AA} \alpha$ exactly corresponds to the variance of the θ_α estimator for different α. The maximizations over α and ρ_0 are due to our requirement that the estimation is oblivious for all α and the algorithm applies to arbitrary reference state ρ_0. The minimization over measurements M guarantees the optimal single-copy measurement is chosen. In particular, adaptivity (i.e. the ability to adjust later measurements based on previous measurement outcomes) provides no advantages, once the high-precision regime is reached. Finally, it is a mathematical property that the function $(I(\rho_0, M)^{-1})_{AA}$ is independent of different choices of dual bases $\{Q_a\}_{a \in A}, \{T_b\}_{b \in A}$ (satisfying Eq. (2.1)), and is solely a function of observables $\{O_i\}_{i=1}^m$. It implies although choosing different dual bases might affect the sample complexity required to distinguish $\rho_{\theta,\varphi}$ from ρ_0, the influence is negligible in the high-precision regime.

To understand the threshold behavior, we explain the derivation of the lower and upper bounds separately. Our lower bound is derived through the distinguishing task (Problem 3') where the goal

Tasks	Oblivious estimation from L^q-ellipsoid (Problems 1' and 2', $p \in [1, \infty]$)	Shadow estimation with p-norm error (Problems 1 and 2, $p \in [1, \infty]$)
Lower bounds for many-versus-one distinguishing (applying to Problems 3')	$\Omega\big(\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)/\varepsilon^2\big)$ ($\varepsilon \leq \eta^{\mathrm{ob}}$, single-copy)	Same as the left column
	$\Omega\big(\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)/(c\varepsilon^2)\big)$ ($\varepsilon \leq \min\{\eta^{\mathrm{ob}}, \eta_c^{\mathrm{ob}}\}$, c-copy)	Same as the left column
Lower bounds for unbiased, bounded estimation	$\Omega\big(\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)/(c\varepsilon^2 \log(1/\varepsilon))\big)$ (c-copy)	$\Omega\big(\Gamma_p(\{O_i\}_{i=1}^m)/(c\varepsilon^2 \log\big(m^{1/p}/\varepsilon\big))\big)$ (c-copy, $p \in [2, \infty]$)
Upper bounds using the two-step method	$O\big(\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)/\varepsilon^2\big)$ ($\varepsilon \leq \overline{\eta}^{\mathrm{ob}}$, single-copy)	$O\big(\log(m)\Gamma_p(\{O_i\}_{i=1}^m)/\varepsilon^2\big)$ ($\varepsilon \leq \overline{\eta}$, single-copy)

Table 1: Summary of results. We use two different methods to derive lower bounds on the sample complexity using c-copy measurements. The first method is detailed in Appendix C where lower bounds and corresponding thresholds are derived for Problem 3' using the learning tree method. The second method is detailed in Appendix D where lower bounds are derived directly for Problems 2 and 2' using the Cramér–Rao method assuming unbiased and bounded estimators. Explicit algorithms are provided in Appendix E using the two-step method where state tomography is first used to find a nearby state and locally optimal estimation is then used to achieve the target precision.

is to distinguish all well-defined $\rho_{\theta,\varphi}$ with $\|\theta\|_p = 3\varepsilon$ from ρ_0. For any $\varepsilon > 0$ without threshold, we show a lower bound equal to

$$\Omega\left(\inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\substack{(\theta,\varphi) \text{ s.t. } \rho_{\theta,\varphi} \succeq 0, \\ \|\theta\|_p = 3\varepsilon}} \Big((\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi)\Big)^{-1}\right). \tag{2.7}$$

Here $(\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi)$ can be interpreted as the Fisher information corresponding to a specific instance of state $\rho_{\theta,\varphi}$ satisfying the well-definedness constraint $\rho_{\theta,\varphi} \succeq 0$ and $\|\theta\|_p = 3\varepsilon$. In order to distinguish successfully all instances from ρ_0, we maximize the function over all possible (θ,φ). In fact, we show the duality between p- and q- norms guarantees

$$\max_{\|\theta\|_p = 3\varepsilon} \big((\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi)\big)^{-1} = \max_{\|\alpha\|_q = 3\varepsilon} \alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha. \tag{2.8}$$

That means in order to prove Eq. (2.4) as a lower bound for the distinguishing task, it is sufficient to remove the constraint $\rho_{\theta,\varphi} \succeq 0$ from Eq. (2.7), which is possible by introducing a threshold. One trick we apply here that allows us to consider only states in $\mathcal{S}_{1/2} := \{\rho | \rho = \sigma/2 + \mathbb{I}/(2d),$ for some density matrix $\sigma\}$ is to observe that $I(\frac{1}{2}\rho_0 + \frac{\mathbb{I}}{2d}, M) \preceq 2I(\rho_0, M)$, which means mixing ρ_0 evenly with a maximally mixed state $\frac{\mathbb{I}}{d}$ does not change our lower bound up to constant. This allows us to derive an explicit formula of a valid threshold η^{ob} as a function of $\{O_i\}_{i=1}^m$ below which Eq. (2.4) is a lower bound.

Conceptually, the requirement of a threshold in the lower bound implies that the distinguishing task may be fundamentally easier to solve than the estimation task when ε is too large, where the ensemble of states $\{\rho_{\theta,\varphi}\}$ satisfying $\|\theta\|_p = 3\varepsilon$ may no longer be a good hypothesis to characterize the difficulty in parameter estimation because $\rho_{\theta,\varphi}$ is not well-defined for too many (θ,φ). For the special case of ∞-norm Pauli estimation, however, the distinguishing task provides a tight lower bound for all $\varepsilon > 0$ (Chen et al., 2024b) (when using single-copy measurements)—this is because restricting (θ, φ) to the set of $(e_i, 0)$ where $e_i \in \mathbb{R}^m$ is a vector that is 1 in the i-th entry and 0 in the others does not change the value of Eq. (2.7). However, it is unknown whether the same property holds for general $p \in [1, \infty]$ and general observables.

The upper bound is derived using a two-step method where state tomography is first used to find a nearby state and locally optimal estimation is then used to achieve the target precision. The two-step method is traditionally used in quantum metrology to show the attainability of the Cramér–Rao bound asymptotically, i.e. when taking the limit $\varepsilon \to 0$ and allowing infinitely many samples. We instead demonstrate the usefulness of the two-step method in a finite-sample regime, that was rarely explored previously, bridging a gap between quantum metrology and quantum learning. One interesting result we manage to show is the locally optimal estimator from quantum metrology is globally optimal (up to a constant factor) and unbiased within a finite-size region containing ρ_0. The bound in Eq. (2.4) is tight when the sample complexity required in the pre-estimation stage for finding the finite-size region is negligible. To bound it, we again restrict our discussion to states in $\mathcal{S}_{1/2}$ using the state mixture trick, and show state tomography that determines a nearby state within $O(1/d)$ in ∞-norm distance away from ρ is sufficient for our purpose. When the target precision ε is too large, the pre-estimation stage can be too costly, making our algorithm suboptimal.

In practice, the assumption of adaptive single-copy measurements can be too restrictive sometimes because entangled measurements across a few copies of states are also feasible on certain experimental platforms (Huang et al., 2022) and they can sometimes bring substantial improvement in sample complexity. For example, for ∞-norm Pauli estimation within the low-precision regime (e.g., $\varepsilon = 0.1$), even two-copy measurements can provide exponential sample complexity reduction compared to single-copy protocols (Chen et al., 2024b; Huang et al., 2021). However, for ∞-norm Pauli estimation within the high-precision regime (e.g. when $\varepsilon = O(1/d^{1/2})$), multi-copy measurements provide no substantial advantages over single-copy measurements unless measurements across an exponentially large number of copies are available. Here we observe a similar phenomenon for general oblivious estimation.

Theorem 2 (Oblivious estimation with c-copy measurements, informal) *Given any $p \in [1, \infty]$, to solve Problem 1', Problem 2' or Problem 3' with (adaptive) c-copy measurement protocols, there exists a threshold on the target precision ε below which*

$$N = \Omega\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{c\varepsilon^2}\right) \tag{2.9}$$

copies of quantum states ρ are necessary.

Theorem 2 implies allowing entangled measurements across c copies of states can at most bring a $O(1/c)$ advantage in the sample complexity in the high-precision regime. When c is not too large, e.g. $c = O(\mathrm{polylog}(d))$, the reduction in sample complexity will also be at most polynomial. We prove the result by first bounding the lowest-order part of the sample complexity for small ε, and then find a threshold on ε below which the higher-order terms are negligible.

Finally, we note that an alternative method to derive lower bounds for Problems 2' (and Problem 2 in the later section) is through the Cramér–Rao bound. It provides a lower bound on the variance of any unbiased estimator given by the inverse of the FIM. Assuming estimator values are always bounded away from true values by at most a constant, the variance of estimation can be directly converted to the oblivious estimation error or the p-norm error with a logarithmic overhead (see Table 1). The bound holds for arbitrary $\varepsilon > 0$ without any threshold, as we restrict our discussion to unbiased, bounded estimators and are no longer considering the distinguishing task (Problem 3').

2.3. Shadow estimation with p-norm error

Above we provide tight bounds for the oblivious estimation and distinguishing tasks. They can be fundamentally easier than the shadow estimation task. Luckily, our two-step algorithm applies also to the general shadow estimation tasks, and the Cramér–Rao method to derive lower bounds also (partly) applies here. As a result, we have the following theorems.

Theorem 3 (Shadow estimation with one-copy measurements, informal) *Given any $p \in [1, \infty]$, there exists an unbiased estimation algorithm using single-copy measurements that solves Problem 1 (and Problem 2) using*

$$N = \tilde{\Theta}\left(\frac{\Gamma_p(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right) \tag{2.10}$$

copies of quantum state ρ when the target precision ε is below a threshold. Here $\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)$ is a positive function of the observables defined by

$$\Gamma_p(\{O_i\}_{i=1}^m) := \inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \left\| \mathrm{diag}\left((I(\rho_0, M)^{-1})_{AA} \right)^{1/2} \right\|_p^2, \tag{2.11}$$

where $\mathrm{diag}(\cdot)$ represents a diagonal matrix whose diagonal entries are those of $(\cdot)$ and $\|\cdot\|_p$ represents the Schatten p-norm of matrices. In particular, for $p \in [2, \infty]$, Eq. (2.10) is also necessary for any $\varepsilon > 0$ when assuming unbiased and bounded estimation.

Theorem 4 (Shadow estimation with c-copy measurements, informal) *Given any $p \in [2, \infty]$, to solve Problem 1 (and Problem 2) with (adaptive) c-copy measurement protocols,*

$$N = \tilde{\Omega}\left(\frac{\Gamma_p(\{O_i\}_{i=1}^m)}{c\varepsilon^2}\right) \tag{2.12}$$

copies of quantum states ρ are necessary, when assuming unbiased and bounded estimation.

To understand why the Cramér–Rao method and the two-step method still apply to the shadow estimation tasks, we note that they provide lower and upper bounds on the variance of estimators which can be converted to and from our p-norm error using standard statistical techniques like median-of-means estimation. One special point is we can only prove lower bound on cases with $p \geq 2$ because when $p \in [1, 2)$, it is not guaranteed a small p-norm error can lead to a small variance of estimation. The result on few-copy measurements follows from a property of FIM, where increasing c, the number of copies of states, can at most increase the FIM by $O(c^2)$.

We note that when $p = \infty$ (which corresponds to the traditional shadow estimation scenario), the two tasks are equivalent in the high-precision regime up to logarithmic overhead.

Corollary 5 (Equivalence between shadow estimation and oblivious estimation, $p = \infty$) *Let $p = \infty$. When ε is below a threshold, the necessary and sufficient sample complexities to solve Problem 1 and to solve Problem 1' using (adaptive) single-copy measurements are both*

$$N = \tilde{\Theta}\left(\frac{\Gamma_\infty(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right), \tag{2.13}$$

where

$$\Gamma_\infty(\{O_i\}_{i=1}^m) = \Gamma_\infty^{\mathrm{ob}}(\{O_i\}_{i=1}^m) = \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{a\in A}(I(\rho_0, M)^{-1})_{aa}. \tag{2.14}$$

The equivalence can be understood from two different perspectives. On one hand, it directly follows from our instance-optimal bounds (Theorem 1 and Theorem 3), and the mathematical property that $\Gamma_\infty(\{O_i\}_{i=1}^m) = \Gamma_\infty^{\mathrm{ob}}(\{O_i\}_{i=1}^m)$. On the other hand, we can use any algorithm that can solve the oblivious estimation task to solve the shadow estimation task by estimating each observable O_i one by one. As a result, if the sample complexity of the oblivious estimation task has a δ dependence of $O(\log(1/\delta))$ where δ is the failure probability—which is satisfied by our algorithm—then the δ dependence becomes $O(\log(m/\delta))$ in shadow estimation by union bound, introducing at most a logarithmic overhead.

Finally, as an example, we show in Appendix F when $\{O_i\}_{i=1}^m$ is the set of all Pauli observables,

$$\Omega(d) \le \Gamma_p^{\mathrm{ob}} \le \begin{cases} O(d\log d)\,, & \text{if } p\in[2,\infty], \\ O\left(d^{\frac{4}{p}-1}\log d\right), & \text{if } p\in[1,2), \end{cases} \qquad \Omega(d)\le\Gamma_2\le O(d^3\log d). \tag{2.15}$$

The scalings of our thresholds (see Table 1) are

$$\eta^{\mathrm{ob}} = \Omega(d^{\frac{2}{p}-2}),\ \eta_c^{\mathrm{ob}} = \begin{cases} \Omega(d^{-\frac{2}{p}+\frac{1}{2}}(\log d)^{-\frac{1}{2}}/c), & \text{if } p\in[1,2), \\ \Omega(d^{-\frac{5}{2}+\frac{4}{p}})(\log d)^{-\frac{1}{2}}/c), & \text{if } p\in[2,\infty], \end{cases} \qquad \overline{\eta}\ge\overline{\eta}^{\mathrm{ob}} = \Omega(d^{-1}). \tag{2.16}$$

The $p < \infty$ case was previously unknown to the best of our knowledge.

Throughout our results, the thresholds can scale inverse polynomially in dimension (i.e., exponentially in n). We emphasize that this threshold is *necessary*: For Pauli observables and infinite norm learning, it was shown that learning all Paulis up to ε error requires $\tilde{\Theta}(\min\{d/\varepsilon^2, 1/\varepsilon^4\})$ samples (Chen et al., 2024b). This necessitates a threshold of $d^{-1/2}$ in order for the $\tilde{\Theta}(d/\varepsilon^2)$ term—which corresponds to our sample complexity bound—to become the dominant term. Note that when the number of observables is $\mathrm{polylog}(d)$, our lower bound threshold may be improved to $1/\mathrm{polylog}(d)$. However, our upper bound threshold is still $O(1/d)$ and it is unclear whether this is a proof artifact or fundamental. In the revision, we will discuss the necessity of exponentially small thresholds. Finally, the high-precision regime is practically relevant for sensing. For example, in the task of imaging of weak incoherent light sources with d telescopes, a d-dimensional single-photon state is detected where each entry of the density matrix scales as $O(1/d)$. The image information can be recovered from the Fourier transforms of these entries (Zernike, 1938; Khabiboulline et al., 2019). In this setting, achieving a learning precision of $O(1/d)$—and hence operating in the high-precision regime—is essential.

3. Technical overview

In this section, we provide an overview of the techniques for all theorems mentioned so far.

3.1. Learning tree method for single-copy measurements

Our lower bounds for the oblivious distinguishing task in Problem 3' (and thus for all remaining problems) throughout this paper exploit and improve the well-established "learning tree" framework gradually developed in a series of work (Aharonov et al., 2022; Bubeck et al., 2020; Chen et al., 2022b, 2023b; Chen and Gong, 2025; Chen et al., 2024b). From a high level, we model any (adaptive) learning protocol as a decision tree, and a choice of the underlying unknown state ρ results in a distribution on the leaves. We then consider the ensemble $\rho_{\theta,\varphi}$ parameterized by dual observable bases $\{Q_a, T_b\}_{a\in A, b\in B}$ with $A = \{1, ..., m\}$ and $B = \{m+1, ..., d^2-1\}$, and $\theta \in \mathbb{R}^m$ and $\varphi \in \mathbb{R}^{d^2-m-1}$. We argue that the resulting distributions on leaves are statistically indistinguishable from the distributions on leaves for ρ_0 unless the depth of the tree is sufficiently large. In this paper, we take the parameterized ensemble $\rho_{\theta,\varphi}$ to be randomly sampled from a probability distribution π. Here, $\rho_{\theta,\varphi}$ is defined to be $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b$, where ρ_0 is the density matrix in the null hypothesis, and θ and φ can be arbitrary vectors such that $\rho_{\theta,\varphi}$ is well-defined and $\|\theta\|_p = 3\varepsilon$. We utilize the learning tree model equipped with martingale analysis (Chen et al., 2023b; Chen and Gong, 2025; Chen et al., 2024b) to obtain a sample complexity lower bound in Eq. (2.7) for (adaptive) single-copy measurement protocols. The detailed derivation is provided in Appendix C.1.

3.2. Duality between distinguishing and estimation, and the threshold on ε

We have shown earlier that the (oblivious) distinguishing task Problem 3(3') is no harder than the (oblivious) estimation task Problem 2(2'). For Problem 2', we can obtain a lower bound of Eq. (2.4) claimed in Theorem 1 for $1/p + 1/q = 1$ using the following quantum metrology argument. We first note that the inverse of the FIM is an asymptotically attainable lower bound on the variance of estimators by the Cramér–Rao bound (Kay, 1993; Lehmann and Casella, 2006). We then show that the scalar $\alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha$ exactly matches the variance of the estimator for θ_α asymptotically with at most poly-logarithmic overhead. We then minimize over vectors α and states ρ_0 as the estimation is oblivious for all α and any reference state ρ_0, and maximize over measurements M as we can use the optimal single-copy measurements. Showing that adaptivity provides no advantages, we reach the lower bound claimed in Theorem 1 for Problem 2'.

To bridge the gap between Eq. (2.7) and Eq. (2.4), we show the duality between p- and q-norm as in Eq. (2.8). However, we note that the (θ, φ) satisfy not only the constraint $\|\theta\|_p \leq 3\varepsilon$, but also the well-definedness of the quantum state $\rho_{\theta,\varphi}$ (i.e. $\rho_{\theta,\varphi} \succeq 0$). Therefore, we still need to remove the constraint of $\rho_{\theta,\varphi} \succeq 0$ to reach the lower bound for Problem 3' claimed in Theorem 1.

An immediate thought is to hope that there is some threshold η such that we can drop $\rho_{\theta,\varphi} \succeq 0$ when $\varepsilon < \eta$. Unfortunately, even when $\{Q_a\}$ is complete and thus $A = \{1, ..., d^2-1\}, B = \varnothing$, there does not exist such a choice of η that applies to all ρ_0 as it can be close to the boundary of the density matrix space. However, we can restrict the range of ρ_0 within $\mathcal{S}_{1/2}$ as the set of $(1/2, 1/2)$ linear mixture state of the maximally mixed state and a full rank state, without decreasing Eq. (2.7). When $\{Q_a\}$ is complete, $I(\rho_0, M)$ is diagonal and we can choose a threshold η to ensure that $\rho_{\theta,\varphi}$ is always well-defined and drop the verbose constraint $\rho_{\theta,\varphi} \succeq 0$. We refer to Theorem 14 for the detailed derivation.

However, when $\{Q_a\}$ is incomplete, $I(\rho_0, M)$ is not naturally block-diagonal. To address this issue, we consider the following linear transformation between $(Q\ T) = (Q_1\ Q_2\ \cdots\ Q_m\ T_{m+1}\ T_{m+2}\ \cdots\ T_{d^2-1})$ and $(Q'\ T') = (Q'_1\ Q'_2\ \cdots\ Q'_m\ T'_{m+1}\ T'_{m+2}\ \cdots\ T'_{d^2-1})$

related by $(Q', T') = (Q + TC_1, TC_2)$, where we $C_1 \in \mathbb{R}^{|B|\times|A|} = \mathbb{R}^{d^2-1-m\times m}$ and $C_2 \in \mathbb{R}^{|B|\times|B|} = \mathbb{R}^{d^2-1-m\times d^2-1-m}$ to represent linear transformations on the matrix blocks. Here, C_1 can be arbitrary and C_2 needs to be invertible. As a result, the corresponding FIMs are related by a corresponding linear transformation. Finally, we only need to take $C_2 = \mathbb{I}$, which means $T' = T$ and $C_1 = -(I(\rho_0, M)_{BB})^{-1} I(\rho_0, M)_{BA}$, where $(I(\rho_0, M)_{BB})^{-1}$ is the pseudoinverse of $I(\rho_0, M)_{BB}$ on its support. The corresponding choice of basis $Q' = Q + TC_1$ makes the corresponding FIM $I(\rho_0, M)'$ block-diagonal. We can choose a threshold η to ensure that $\rho_{\theta,\varphi}$ is always well-defined and drop the verbose constraint $\rho_{\theta,\varphi} \succeq 0$. We refer to Theorem 16 for the detailed derivation.

3.3. Learning tree method for few-copy measurements

We now consider generalizing the above lower bound argument to c-copy measurements. For technical simplicity, we slightly modified Problem 3' and allow the reference state ρ_0 to be randomly chosen. Fix T in the dual basis. Let $\widetilde{\mathcal{D}}$ denote the set of all probability distributions of $(\rho_0, \tilde{Q}, \theta, \varphi)$ where $\tilde{Q} = Q + TC_1$ for arbitrary C_1 and some fixed Q. We can then use the minimax theorem (Sion, 1958) to change the sequence of minimization over $\tilde{D}$ and maximization over M. An important property implied by the above argument is that, assume $\pi^\star$ is a nearly optimal distribution for some fixed T and $\Lambda_{c_2}\pi(\rho_0, \tilde{Q}, \theta, \varphi) := c_2\pi(\rho_0, \tilde{Q}, \theta, c_2\varphi)$, $\Lambda_{c_2}\pi^\star$ for any $c_2 \geq 1$ is also a near-optimal distribution with a constant overhead independent of c_2.

Using the learning tree framework, the sample complexity bound for Problem 3' using c-copy measurements is $\Omega(c/\delta_{\mathcal{M}_c}(O))$, where $\delta_{\mathcal{M}_c}(O) = \sup_{M\in\mathcal{M}_c} \inf_{\rho_0} \min_{(\theta,\varphi)} \chi^2_M\big(\rho_{\theta,\varphi}^{\otimes c} \| \rho_0^{\otimes c}\big)$. We then relax ρ_0 to be randomly chosen, and decouple $\sqrt{\delta_{\mathcal{M}_c}}(O)$ into the first-order term that depends on one copy and the higher-order terms that depend on more than one copy. We first argue that the first-order term exactly scales as c^2 times the single-copy $\delta_{\mathcal{M}}(O)$. For the higher-order terms, we can apply the Λ_{c_2} transformation to infinitely squeeze in the domain of φ and collapse the distribution into a delta function. As a result, we can reduce the upper bound on the higher-order terms into an optimization problem only over (distributions of) ρ_0, M, and θ. We then compute the threshold on ε such that the higher-order terms are negligible compared to the first-order term. The detailed derivation is provided in Theorem 18.

3.4. Cramér–Rao method

The Cramér–Rao bound (Kay, 1993; Lehmann and Casella, 2006) states that $V(\rho_\theta, M, \hat{\theta}) \succeq I(\rho_\theta, M)^{-1}$, where $V(\rho_\theta, M, \hat{\theta})$ is the mean square error matrix (MSEM) of an unbiased estimator $\hat{\theta}$ when performing a single-copy measurement M on a parameterized state ρ_θ. In our case, we assume unbiased and bounded estimators when estimating $\rho_{\theta,\varphi}$. To prove the lower bound on the shadow estimation problem (Theorem 20 for Problem 2), we apply the following techniques: (1) we prove a lower bound on the number of repeated single-copy measurements needed to achieve a certain p-average root mean square error (RMSE), which is essentially the p-norm of the square root of the diagonal elements of the MSEM, directly using the CR bound; (2) we show that when $p \geq 2$, if an estimator has a small p-norm error, then it can also achieve a small p-average RMSE with a logarithmic overhead; (3) we prove that adaptivity in measurements does not decrease the value of the lower bound using the chain rule of Fisher information. Here we require the estimator to be bounded in order to rule out the possibility that the estimator takes very large values on a negligible support which may potentially contribute non-negligibly to the p-average RMSE. To prove the lower bound on the oblivious estimation problem (Theorem 19 for Problem 2'), the steps are very similar, except that in the second step the conversion from additive error to RMSE applies to all $p \in [1, \infty]$. Finally, in order to prove the lower bounds with c-copy measurements (Theorem 21), we use a previously used technique in Theorem 18, which shows $I(\rho_0^{\otimes c}, M)^{1/2} \preceq \sum_{i=1}^{c} I(\rho_0, G^{[i]})^{1/2}$, where M is a c-copy measurement and $G^{[i]}$ is some single-copy measurement that depends on M, ρ_0 and i. It then implies for any c-copy measurement M, there is some corresponding single-copy measurement G such that $I(\rho_0^{\otimes c}, M)^{1/2} \preceq cI(\rho_0, G)^{1/2}$. As a result, the c-copy lower bounds are at most a factor of $1/c$ smaller than the single-copy lower bounds.

3.5. Two-step method: from local to global estimation

The two-step method is traditionally used in quantum metrology to show the attainability of the CR bound (Barndorff-Nielsen and Gill, 2000; Hayashi, 2011; Yang et al., 2019). In those cases, one starts from a locally unbiased estimator, whose expected value equals the true parameter value at (and infinitesimally around) a specific point, that achieves the CR bound at that point. The goal is to show that, without any prior knowledge of the location of the point, one can still achieve the CR bound asymptotically. If N is the total number of samples, the two-step method first uses a negligible number of samples, e.g. $\sqrt{N}$, to obtain a coarse estimate of the parameter, and then applies the locally unbiased estimator defined at the coarse estimation point. The attainability of the CR bound as $N \to \infty$ can be shown when suitable bounds on the convergence rate are available.

In the context of quantum learning, we aim to obtain the sample complexity needed for a finite target precision. The main challenge is to apply the two-step method in a non-asymptotic manner, and the advantage we can leverage is that we allow a logarithmic-factor discrepancy, whereas in quantum metrology the CR bound must be strictly attained. To find algorithms that saturate our lower bounds up to a logarithmic factor (Theorem 25 for Problem 2, and Theorem 26 for Problem 2'), we first show that within a region $\mathscr{N}(\rho_0)$ around a specific point ρ_0, the locally unbiased estimator at ρ_0 achieves the CR bound up to a factor of two for any state in $\mathscr{N}(\rho_0)$. The key observation is that $\rho_{\theta,\varphi}$ is linear in θ, such that any locally unbiased estimator of θ must also be a globally unbiased estimator. To find a coarse estimate $\hat{\rho}_0$ of ρ such that $\rho \in \mathscr{N}(\hat{\rho}_0)$, we show that it is sufficient to apply state tomography with a $O(1/d)$ inaccuracy in operator norm. Here we consider only states whose minimum eigenvalues are above $1/2d$, as we can always add artificial depolarizing noise into the system. Finally, by allowing a logarithmic overhead, we turn our estimator, which has a small p-average RMSE, into an estimator with a small p-norm error using the median-of-means estimation. In order to attain the lower bounds, we need ε to be small enough such that the coarse estimation step takes a negligible number of samples.

4. Discussion and outlook

In this work, we give an instance-optimal characterization of high-precision shadow tomography—generalized to p-norm error—by expressing the fundamental sample complexity through Fisher-information-based quantities Γ_p and Γ_p^{ob}, and showing the $\tilde{\Theta}(\Gamma_p/\varepsilon^2)$ scaling below explicit thresholds with matching upper and lower bounds in the relevant regimes. Our results also clarify the role of experimentally feasible entanglement: in the high-precision regime, c-copy measurements can improve the leading term by at most a factor $O(1/c)$, indicating substantial asymptotic gains would require access to measurements across much larger numbers of copies. Conceptually, the work bridges quantum learning and quantum metrology by demonstrating that a finite-sample two-step procedure—coarse localization via tomography followed by locally optimal estimation—can be near-globally optimal on a finite neighborhood rather than only asymptotically. Below, we mention some concrete open questions closely related to the current work.

Linearly-dependent observables. It would be interesting to consider linearly dependent observables. We can consider (wlog) estimating $o \in \mathbb{R}^m$ through $o = W\theta$, where $W \in \mathbb{R}^{m\times r}$ is isometric and $\theta \in \mathbb{R}^r$ and replace θ and α with $W\theta$ and $W\alpha$ in the definitions of Problems. We expect our two-step method and Cramér-Rao method to still apply because the mean square matrix wrt o is equal to that wrt θ under a congruent transformation. The learning tree method, however, will be problematic in general, because the new Problem 3' no longer reduces to Problem 2'. This is roughly because p-norm is not invariant under isometric transformation, except when $p = 2$, a special case where we expect our learning tree method to directly apply.

Closed-form thresholds. We have derived high-precision thresholds matching the lower and upper bounds up to logarithmic factors. For incomplete observables, we give a construction via a basis transform that yields a valid threshold. It is interesting to explore if we can turn these constructive thresholds into genuinely closed-form and ideally tight characterizations expressed directly in terms of the observables.

Entangled measurements on a large number of samples. We have investigated protocols with joint measurements across c copies at $c = \text{polylog}(d)$ in the high-precision regime. However, it remains open to study protocols with entangled measurements on a large number of samples (e.g. $c = \text{poly}(d)$).

Conditions for a polylog$(d)^{-1}$ threshold. As we have explicitly computed for the case of Pauli observables, the threshold for the high precision regime scales as $\text{poly}(d)^{-1}$. It is natural to ask if we can identify the conditions on the set of observables and the index p such that the threshold for the high precision regime scales as polylog$(d)^{-1}$.

Acknowledgments

The authors thanks Sitan Chen, Yunchao Liu and Yuxiang Yang for valuable discussions and feedback. S.C. acknowledges support from the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-2317110). S.Z. acknowledges funding provided by Perimeter Institute for Theoretical Physics, a research institute supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. S.C. and S.Z. also acknowledge support from the Kavli Institute for Theoretical Physics (NSF Grant PHY-2309135), where part of this work was completed. W.G. acknowledges support by the Von Neumann Award from Harvard Computer Science and NSF Grant CCF-2430375.

Note: After our paper appeared, a concurrent and related work Kwon et al. (2026) was posted on arXiv, which also connects quantum metrology to quantum learning theory. It obtains sample complexity bounds for maximum-likelihood estimators using different approaches from ours.

References

Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 325–338, 2018.

Scott Aaronson and Guy N Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333, 2019.

Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. In *Advances in neural information processing systems*, volume 31, 2018.

Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(887):1–9, 2022.

Francesco Albarelli, Jamie F Friel, and Animesh Datta. Evaluating the holevo Cramér-Rao bound for multiparameter quantum metrology. *Physical Review Letters*, 123(20):200503, 2019a.

Francesco Albarelli, Mankei Tsang, and Animesh Datta. Upper bounds on the holevo cramér-rao bound for multiparameter quantum parametric and semiparametric estimation. *arXiv:1911.11036*, 2019b.

Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Reviews Physics*, pages 1–11, 2023.

Srinivasan Arunachalam and Ronald De Wolf. Guest column: A survey of quantum learning theory. *ACM Sigact News*, 48(2):41–67, 2017.

Konrad Banaszek, Marcus Cramer, and David Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, 2013.

O E Barndorff-Nielsen and R D Gill. Fisher information in quantum statistics. *J. Phys. A: Math. Gen.*, 33(24):4481–4490, jun 2000.

Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12 (4):043034, 2010.

Fernando GSL Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M Svore, and Xiaodi Wu. Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

Samuel L. Braunstein and Carlton M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72(22):3439–3443, May 1994.

Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P Lanyon, Peter Zoller, Rainer Blatt, and Christian F Roos. Probing Rényi entanglement entropy via randomized measurements. *Science*, 364(6437):260–263, 2019. doi: 10.1126/science.aau4963.

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.

Sebastien Bubeck, Sitan Chen, and Jerry Li. Entanglement is necessary for optimal quantum property testing. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 692–703. IEEE, 2020.

Angelo Carollo, Bernardo Spagnolo, Alexander A Dubkov, and Davide Valenti. On quantumness in multi-parameter quantum estimation. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(9):094010, 2019.

Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. Quantum advantages for Pauli channel estimation. *Physical Review A*, 105(3):032435, 2022a.

Senrui Chen, Changhun Oh, Sisi Zhou, Hsin-Yuan Huang, and Liang Jiang. Tight bounds on Pauli channel learning without entanglement. *Physical Review Letters*, 132(18):180805, 2024a.

Sitan Chen and Jordan Cotler. Lecture 6: Tomography ii: Single-copy measurements (operator norm). Lecture notes for Harvard Physics 272 / CS 2233: Quantum Learning Theory (Fall 2025), September 2025. URL https://harvard-quantum-learning.github.io/lecture6.pdf. Accessed: 2026-01-16.

Sitan Chen and Weiyuan Gong. Efficient pauli channel estimation with logarithmic quantum memory. *PRX Quantum*, 6(2):020323, 2025. doi: 10.1103/PRXQuantum.6.020323.

Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. A hierarchy for replica quantum advantage. *arXiv:2111.05874*, 2021.

Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585. IEEE, 2022b.

Sitan Chen, Jerry Li, Brice Huang, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1205–1213. IEEE, 2022c.

Sitan Chen, Jerry Li, and Ryan O'Donnell. Toward instance-optimal state certification with incoherent measurements. In *Conference on Learning Theory*, pages 2541–2596. PMLR, 2022d.

Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of NISQ. *Nature Communications*, 14(1):6001, 2023a.

Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of nisq. *Nat. Commun.*, 14(1):6001, 2023b.

Sitan Chen, Weiyuan Gong, and Qi Ye. Optimal tradeoffs for estimating Pauli observables. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1086–1105, 2024b. doi: 10.1109/FOCS61266.2024.00072.

Sitan Chen, Jerry Li, and Allen Liu. Optimal high-precision shadow estimation. *arXiv:2407.13874*, 2024c.

Sitan Chen, Jerry Li, and Allen Liu. An optimal tradeoff between entanglement and copy complexity for state tomography. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1331–1342, 2024d.

Giulio Chiribella, Giacomo Mauro D'Ariano, and Dirk Schlingemann. How continuous quantum measurements in finite dimensions are actually discrete. *Phys. Rev. Lett.*, 98(19):190403, 2007. doi: 10.1103/PhysRevLett.98.190403.

Alexander M. Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T. Hann, Michael J. Kastoryano, Emil T. Khabiboulline, Aleksander Kubica, Grant Salton, Samson Wang, and Fernando G. S. L. Brandão. *Quantum Algorithms: A Survey of Applications and End-to-end Complexities*. Cambridge University Press, 2025. ISBN 9781009639644. doi: 10.1017/9781009639651.

Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80(1):012304, 2009. doi: 10.1103/PhysRevA.80.012304.

Christian L Degen, Friedemann Reinhard, and Paola Cappellaro. Quantum sensing. *Reviews of modern physics*, 89(3):035002, 2017. doi: 10.1103/RevModPhys.89.035002.

Rafał Demkowicz-Dobrzański, Wojciech Górecki, and Mădălin Guţă. Multi-parameter estimation beyond quantum fisher information. *Journal of Physics A: Mathematical and Theoretical*, 53 (36):363001, 2020.

Andreas Elben, Steven T Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1): 9–24, 2023. doi: 10.1038/s42254-022-00535-2.

Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347, 2005. doi: 10.1088/1464-4266/7/10/021.

Omar Fawzi, Nicolas Flammarion, Aurélien Garivier, and Aadil Oufkir. Quantum channel certification with incoherent strategies. In *COLT 23-36th Annual Conference on Learning Theory*, pages 1–58, 2023.

James W Gardner, Tuvia Gefen, Simon A Haine, Joseph J Hope, and Yanbei Chen. Achieving the fundamental quantum limit of linear waveform estimation. *Physical Review Letters*, 132(13): 130801, 2024.

Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nature photonics*, 5(4):222–229, 2011. doi: 10.1038/nphoton.2011.35.

Weiyuan Gong and Scott Aaronson. Learning distributions over quantum measurement outcomes. In *International Conference on Machine Learning*, pages 11598–11613. PMLR, 2023.

Wojciech Górecki, Sisi Zhou, Liang Jiang, and Rafał Demkowicz-Dobrzański. Optimal probes and error-correction schemes in multi-parameter quantum metrology. *Quantum*, 4:288, 2020.

David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15):150401, 2010.

M Guţă, J Kahn, R Kueng, and J A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020. ISSN 1751-8113, 1751-8121. doi: 10.1088/1751-8121/ab8111.

Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, pages 913–925, 2016.

Robin Harper, Steven T Flammia, and Joel J Wallman. Efficient learning of quantum noise. *Nature Physics*, 16(12):1184–1188, 2020.

Aram W Harrow. The church of the symmetric subspace. *arXiv:1308.6595*, 2013.

Akel Hashim, Long B Nguyen, Noah Goss, Brian Marinelli, Ravi K Naik, Trevor Chistolini, Jordan Hines, JP Marceaux, Yosep Kim, Pranav Gokhale, et al. A practical introduction to benchmarking and characterization of quantum computers. *arXiv:2408.12064*, 2024.

Masahito Hayashi. Comparison between the Cramer-Rao and the mini-max approaches in quantum channel estimation. *Commun. Math. Phys.*, 304(3):689–709, 2011.

Masahito Hayashi and Yingkai Ouyang. Tight Cramér-Rao type bounds for multiparameter quantum metrology through conic programming. *Quantum*, 7:1094, 2023.

Carl W Helstrom. Minimum mean-squared error of estimates in quantum statistics. *Physics Letters A*, 25(2):101–102, 1967. doi: 10.1016/0375-9601(67)90366-0.

Carl W Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1: 231–252, 1969. doi: 10.1007/BF01007479.

Carl Wilhelm Helstrom. The minimum variance of estimates in quantum signal detection. *IEEE Trans. Inf. Theory*, 14(2):234–242, 1968. doi: 10.1109/TIT.1968.1054108.

Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1963.10500830.

Alexander S Holevo. *Probabilistic and statistical aspects of quantum theory*, volume 1. Springer Science & Business Media, 2011.

Zdenek Hradil. Quantum state estimation. *Physical Review A*, 55(3):R1561, 1997.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, October 2020. ISSN 1745-2473, 1745-2481. doi: 10.1038/s41567-020-0932-7.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021.

Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, et al. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022.

Jonas Kahn and Mădălin Guţă. Local asymptotic normality for finite dimensional quantum systems. *Communications in Mathematical Physics*, 289(2):597–652, 2009.

Steven M Kay. *Fundamentals of statistical signal processing: Volume I Estimation theory*. Prentice-Hall, Inc., 1993.

Emil T Khabiboulline, Johannes Borregaard, Kristiaan De Greve, and Mikhail D Lukin. Quantum-assisted telescope arrays. *Physical review A*, 100(2):022316, 2019.

Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. Triply efficient shadow tomography. *arXiv:2404.19211*, 2024.

Emanuel Knill, Dietrich Leibfried, Rolf Reichle, Joe Britton, R Brad Blakestad, John D Jost, Chris Langer, Roee Ozeri, Signe Seidelin, and David J Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77(1):012307, 2008. doi: 10.1103/PhysRevA.77.012307.

Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

Hyukgun Kwon, Seok Hyung Lie, and Liang Jiang. Universal sample complexity bounds in quantum learning theory via fisher information matrix. *arXiv:2602.21510*, 2026.

Erich L Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

Nan Li, Christopher Ferrie, Jonathan A Gross, Amir Kalev, and Carlton M Caves. Fisher-symmetric informationally complete measurements for pure states. *Physical Review Letters*, 116(18): 180402, 2016.

Keiji Matsumoto. A new approach to the Cramér-Rao-type bound of the pure-state model. *Journal of Physics A: Mathematical and General*, 35(13):3111, 2002.

Jan Nöller, Viet T Tran, Mariami Gachechiladze, and Richard Kueng. An infinite hierarchy of multi-copy quantum learning tasks. *arXiv:2510.08070*, 2025.

Ryan O'Donnell and Chirag Wadhwa. Instance-optimal quantum state certification with entangled measurements. *arXiv:2507.06010*, 2025.

Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, pages 899–912, 2016.

Matteo GA Paris. Quantum estimation for quantum technology. *Int. J. Quantum Inf.*, 7(supp01): 125–137, 2009.

Angelos Pelecanos, Jack Spilecki, and John Wright. The debiased keyl's algorithm: a new unbiased estimator for full state tomography. *arXiv:2510.07788*, 2025.

Luca Pezze, Augusto Smerzi, Markus K Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Reviews of Modern Physics*, 90(3): 035005, 2018. doi: 10.1103/RevModPhys.90.035005.

Stefano Pirandola, B Roy Bardhan, Tobias Gehring, Christian Weedbrook, and Seth Lloyd. Advances in photonic quantum sensing. *Nature Photonics*, 12(12):724–733, 2018. doi: 10.1038/s41566-018-0301-6.

Herbert Robbins. A remark on Stirling's formula. *The American mathematical monthly*, 62(1): 26–29, 1955.

Alireza Seif, Senrui Chen, Swarnadeep Majumder, Haoran Liao, Derek S Wang, Moein Malekakhlagh, Ali Javadi-Abhari, Liang Jiang, and Zlatko K Minev. Entanglement-enhanced learning of quantum processes at scale. *arXiv:2408.03376*, 2024.

Jasminder S Sidhu, Yingkai Ouyang, Earl T Campbell, and Pieter Kok. Tight bounds on the simultaneous estimation of incompatible parameters. *Physical Review X*, 11(1):011028, 2021.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

Jun Suzuki, Yuxiang Yang, and Masahito Hayashi. Quantum state estimation with nuisance parameters. *Journal of Physics A: Mathematical and Theoretical*, 53(45):453001, 2020.

Mankei Tsang, Francesco Albarelli, and Animesh Datta. Quantum semiparametric estimation. *Physical Review X*, 10(3):031023, 2020.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

C Vargas, L Pereira, and A Delgado. Near-optimal pure state estimation with adaptive fisher-symmetric measurements. *arXiv:2412.04555*, 2024.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 1 edition, 2018. ISBN 978-1-108-23159-6 978-1-108-41519-4. doi: 10.1017/9781108231596.

Adam Bene Watts and John Bostanci. Quantum event learning and gentle random measurements. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024.

Koichi Yamagata, Akio Fujiwara, and Richard D. Gill. Quantum local asymptotic normality based on a new quantum likelihood ratio. *The Annals of Statistics*, 41(4):2197 – 2217, 2013.

Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. Attaining the ultimate precision limit in quantum state estimation. *Communications in Mathematical Physics*, 368(1):223–293, 2019.

Qi Ye, Zhenhuan Liu, and Dong-Ling Deng. Exponential advantage from one more replica in estimating nonlinear properties of quantum states. *arXiv:2509.24000*, 2025.

Bin Yu. Assouad, Fano, and Le Cam. *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435, 1997. URL `https://www.stat.berkeley.edu/~binyu/summer08/yu.assoua.pdf`.

Frederik Zernike. The concept of degree of coherence and its application to optical problems. *Physica*, 5(8):785–795, 1938.

Sisi Zhou and Senrui Chen. Randomized measurements for multiparameter quantum metrology. *PRX Quantum*, 7(1):010314, 2026.

Huangjun Zhu and Masahito Hayashi. Universally Fisher-symmetric informationally complete measurements. *Physical Review Letters*, 120(3):030404, 2018.

Appendix A. Related works

Technically, the lower bounds proved in our work rely on the tools that improved upon the refined learning tree framework from a line of aforementioned works (Bubeck et al., 2020; Chen et al., 2022b, 2023a; Chen and Gong, 2025; Chen et al., 2024b). Here, we mention some other relevant works.

Shadow tomography. The most standard task in quantum learning is quantum state tomography, which completely recovers the density matrix of an unknown quantum state to high accuracy in trace norm or fidelity (Banaszek et al., 2013; Blume-Kohout, 2010; Gross et al., 2010; Hradil, 1997). Unfortunately, quantum state tomography suffers from an unavoidable exponential scaling on the system size in sample complexity (Haah et al., 2016; O'Donnell and Wright, 2016).

To circumvent this exponential barrier, a widely studied alternative task is shadow tomography (Aaronson, 2018), the goal of which is to estimate the expectation values of a set of m observables or measurements (up to certain additive, i.e. ∞-norm, error). A line of works has proposed sample-efficient algorithms using $\mathrm{poly}(\log m, n, 1/\varepsilon)$ copies of unknown quantum states and highly entangled measurements (Aaronson, 2018; Aaronson et al., 2018; Aaronson and Rothblum, 2019; Bădescu and O'Donnell, 2021; Brandão et al., 2019; Gong and Aaronson, 2023; Watts and Bostanci, 2024). Up to date, the gap between the best known sample complexity upper bound scaling as $O((\log^2 m)n/\varepsilon^4)$ (Bădescu and O'Donnell, 2021) and the well-known lower bound of $\Omega(\log m/\varepsilon^2)$ remains an open question. More recently, an optimal shadow tomography protocol with sample complexity $O(\log m/\varepsilon^2)$ is proposed in the high precision regime of $\varepsilon \lesssim d^{-12}$ (Chen et al., 2024c), with a very recent breakthrough improving this threshold to $\varepsilon \lesssim d^{-1}$ (Pelecanos et al., 2025). From the lower bound perspective, it is known that an $\Omega(\min\{2^n/\varepsilon^2, 1/\varepsilon^4\})$ scaling is necessary if one can only perform joint measurements on a restricted number of copies even for estimating all Pauli observables (Chen et al., 2024b).

All above protocols and limitations assume access to highly joint measurements, rendering them impractical to implement on near-term devices. In settings of shadow tomography with single-copy measurements, the classical shadows protocol of Huang, Kueng, and Preskill (Huang et al., 2020) requires $O(2^n \log m/\varepsilon^2)$, which is proved to be optimal up to log factors when $m \geq 2^n$ (Chen et al., 2022b). The algorithm is designed to use random basis measurements to produce a classical, unbiased estimator of the state, which is then used to predict arbitrary observables. Following this random scheme (Elben et al., 2023), more randomized unitaries and measurements are widely applied in quantum device benchmarking (Knill et al., 2008; Dankert et al., 2009; Emerson et al., 2005), and quantum learning and tomography (Elben et al., 2023; Brydges et al., 2019). Note that with few-copy measurements, efficient sample complexity can be achieved for Pauli estimation when the target precision is low, but not in the high-precision regime (Huang et al., 2021; Chen et al., 2024b; King et al., 2024).

Our work also discussed protocols with single-copy and few-copy measurements, with a pioneering focus on instance optimality and general p-norm error, which can be regarded as part of a larger body of recent results exploring how local and joint measurements for various quantum learning tasks affect the underlying statistical complexity (Aharonov et al., 2022; Bubeck et al., 2020; Chen et al., 2022a, 2024a; Seif et al., 2024; Chen et al., 2021; Ye et al., 2025; Nöller et al., 2025; Chen et al., 2022c; O'Donnell and Wadhwa, 2025; Chen et al., 2022d; Fawzi et al., 2023; Huang et al., 2022, 2021; Chen et al., 2024d). We refer to the survey (Anshu and Arunachalam, 2023) for a more thorough overview along and beyond this line of work.

Quantum metrology. Quantum metrology studies optimal measurements for estimating parameters of quantum systems with locally unbiased estimators (Giovannetti et al., 2011; Degen et al., 2017; Pezze et al., 2018; Pirandola et al., 2018). For estimating quantum states parametrized by a single parameter, the quantum Cramér–Rao bound characterizes the ultimate precision limit of estimating one parameter in quantum states through the quantum Fisher information matrix (Helstrom, 1967, 1968, 1969; Holevo, 2011; Braunstein and Caves, 1994; Barndorff-Nielsen and Gill, 2000; Paris, 2009), which is the classical Fisher information matrix maximized over all possible quantum measurements on quantum states. However, when estimating quantum states parametrized by multiple parameters, there are cases where the quantum Cramér–Rao bound and the quantum Fisher information matrix are not achievable even asymptotically.

To address this measurement incompatibility issue, one solution is to minimize the weighted sum of estimation variances for a fixed given cost matrix instead of optimizing the estimation variances for all parameters. In Ref. (Holevo, 2011), a well-known lower bound, known as the Holevo Cramér–Rao bound, is proposed for the weight sum and has been further studied by a line of works (Albarelli et al., 2019a; Górecki et al., 2020; Tsang et al., 2020; Sidhu et al., 2021; Hayashi and Ouyang, 2023; Gardner et al., 2024). In the setting of asymptotically many copies where one can perform joint measurements on infinitely many copies of quantum states, the Holevo Cramér–Rao bound is proved to be tight (Kahn and Guţă, 2009; Yamagata et al., 2013; Yang et al., 2019). When restricted to single-copy measurements on single copies of states, the Holevo Cramér–Rao bound is attainable for pure states (Matsumoto, 2002). Moreover, the Holevo Cramér–Rao bound is shown to be stronger than the Quantum Cramér–Rao bound by at most a factor of 2 (Albarelli et al., 2019b; Carollo et al., 2019; Demkowicz-Dobrzański et al., 2020). Unlike these results, which assume an infinite number of samples and use the (weighted) estimation variance as the metric, our work lies in the finite-sample regime and applies to a different metric.

Connections between shadow tomography and quantum learning. It is natural to connect multi-parameter quantum metrology with (local) quantum state and shadow tomography. A line of work in multi-parameter estimation focuses on local state (shadow) tomography, where optimal measurements, known as Fisher-symmetric measurements, were found and studied for uniformly estimating all parameters in pure states (Li et al., 2016; Zhu and Hayashi, 2018; Vargas et al., 2024). Conceptually, most closely related to the present work are the aforementioned works from Pelecanos, Spilecki, and Wright (Pelecanos et al., 2025), and from Chen and Zhou (Zhou and Chen, 2026), which explore the connections between estimators in shadow tomography and quantum metrology; they apply learning tools to metrology, while we apply a metrological approach to learning. In Ref. (Zhou and Chen, 2026), a protocol using randomized and single-copy measurements is proposed as the locally unbiased estimator for quantum metrology, the mean square error matrix of which is further proved to be within a factor of 4 of being optimal for pure states. They also generalized their results to hold for low-rank states. Later, using the debiased Keyl's estimator, which requires joint measurements, Ref. (Pelecanos et al., 2025) obtains a locally unbiased estimator for quantum metrology with mean square error matrix at most a factor of 2 from being optimal.

Appendix B. Preliminaries

In this section, we recap the basic concepts and results required throughout this paper. We use $\|A\|_p$ to represent the Schatten p-norm of matrix A, $\|A\|_\infty$ to represent the operator norm of A, and

$\|v\|_p$ to represent the L_p norm of the vector v. We also use $\tilde{O}$ and $\tilde{\Theta}$ to hide the poly-logarithmic dependence (on m and $1/\varepsilon$) in big-O notations. We will use $[c]$ to denote the set $\{1, 2, ..., c\}$. When we say "with high probability" without specification, we mean with probability at least $2/3$. We use $\mathbb{I}[\cdot]$ as the indicator function. We use $\succeq$ and $\preceq$ to denote partial orders on positive semidefinite matrices. Given two distributions p and q, the total variation distance between p and q is defined to be $d_{TV}(p, q) := \frac{1}{2}\sum_i |p_i - q_i|$, and the χ^2-divergence between p and q is defined to be $\chi^2(q\|p) := \sum_i p_i \left(\frac{q_i}{p_i} - 1\right)^2$.

B.1. Basic results in quantum information

We first introduce some standard definitions and calculations in quantum information. We consider quantum states in d-dimensional Hilbert spaces represented as a positive semi-definite matrices $\rho \in \mathbb{C}^{d\times d}$ with $\mathrm{tr}(\rho) = 1$. In particular, $d = 2^n$ for n-qubit quantum states. When ρ is rank-1 and thus $\mathrm{tr}(\rho^2) = 1$, it is called a pure state and is denoted as $|\psi\rangle$ or $|\phi\rangle$ throughout this paper. An n-qubit observable $O \in \mathbb{C}^{d\times d}$ is a Hermitian matrix. For an n-qubit quantum state ρ or observable O and a subset $S \in [n]$, we use $\mathrm{tr}_S(\rho)$ or $\mathrm{tr}_S(O)$ to denote the remaining state or observables after tracing out the qubits in S. We denote by $\mathbb{I}$ the identity operator.

For simplicity, we denote by $\mathcal{S}$ the set of all density matrices, $\mathcal{S}^\circ$ the set of all full-rank density matrices, and $\mathcal{S}_{1/2}$ the set of all mixed states that can be written as a linear combination of $\mathbb{I}/d$ and a mixed state $\rho \in \mathcal{S}$ of equal weight $1/2$. We will also consider parameterized quantum state $\rho_{\theta,\varphi}$ with $\rho_{\mathbf{0},\mathbf{0}} = \rho_0$. We denote $\mathscr{D}(\rho_0)$ to be the set of (θ, φ) such that $\rho_{\theta,\varphi}$ is well-defined. We will also consider linear combination of observables. Given a weight vector α and a set of observables $\{O_i\}_i$, we will also denote the linear combination as $O_\alpha = \sum_i \alpha_i O_i$.

Quantum measurements. A general quantum measurement is represented as positive operator-valued measures (POVMs). An n-qubit POVM is represented a set of positive-semidefinite matrices $\{F_s\}_s$ with $\sum_s F_s = \mathbb{I}$ and each F_s a POVM element corresponding to measurement outcome s. When measuring a quantum state ρ with POVM $\{F_s\}_s$, the probability of observing outcome s is given by $\mathrm{tr}(F_s\rho)$.

More generally, POVM is defined to be a mapping of a measurable set $S \subseteq \Omega$, the outcome space, to non-negative Hermitian operators, such that $F(\varnothing) = 0$, $F(\Omega) = \mathbb{I}$, $F(\cup_{i=1}^\infty S_i) = \sum_{i=1}^\infty F(S_i)$ and $\mathrm{tr}(\rho F(S))$ is the probability of obtaining measurement outcomes $\in S$. In this paper, without specification, we will assume without loss of generality $\Omega = \{s\}_s$ is discrete and denotes $F(\{s\}) =: F_s$ as above.

Throughout this paper, we consider quantum measurements on multiple copies (replicas) of d-dimensional quantum states. We denote $\mathcal{M}$ as set of all qudit (single-copy) POVMs. We denote $\mathcal{M}^{[d^2]}$ all qudit (single-copy) POVMs with at most d^2 outcomes. For any integer $c > 1$, we denote $\mathcal{M}_c$ as the set of c-copy POVMs on c qudits. We will also consider learning protocols represented by $\mathcal{M}_{c,N}$, which contains the set of POVMs on N copies of qudit states, which can be decomposed into N/c (possibly adaptive) c-copy measurements.

We will also consider Pauli observables in this paper. We define n-qubit Pauli group $\mathcal{P}_n = \{\mathbb{I}_2, X, Y, Z\}^{\otimes n}$ to be the set of n-qubit Pauli observables, where

$$\mathbb{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{B.1}$$

are single-qubit Pauli operators.

Haar random unitaries. A common tool to prove bounds on the sample complexity of learning problems is to consider random instances. Haar random unitaries is the most common random unitary ensemble. The Haar measure μ on the unitary group $U(d)$ is the unique probability measure that is invariant under left- and right-multiplication

$$\mathbb{E}_{U\sim\mu} f(UV) = \mathbb{E}_{U\sim\mu} f(VU) = \mathbb{E}_{U\sim\mu} f(U), \tag{B.2}$$

for any unitary $V \in U(d)$ and function $f(\cdot)$. We can also define a unique rotation invariant measure on states by $U|\psi\rangle$ with $U \sim \mu$ and an arbitrary state $|\psi\rangle$. We will also write $\psi \sim \mu$ throughout this paper.

We will need to explicitly compute expectation values over the Haar measure. A key subroutine is the following folklore formula (see e.g. Ref. (Harrow, 2013)):

$$\mathbb{E}_{\psi\sim\mu}\left[|\psi\rangle\langle\psi|^{\otimes k}\right] = \frac{\Pi_k}{\binom{d+k-1}{k}} = \frac{1}{d(d+1)...(d+k-1)}\sum_{\pi\in S_k}\pi^d, \tag{B.3}$$

where Π_k denotes the projector onto the symmetric subspace $\mathrm{Sym}^k(\mathbb{C}^d)$, S_k is the set of permutation over k elements and π^d acts on $(\mathbb{C}^d)^{\otimes k}$ by

$$\pi^d|i_1,\dots,i_k\rangle = |i_{\pi^{-1}(1)},\dots,i_{\pi^{-1}(k)}\rangle. \tag{B.4}$$

B.2. Tree representations and Le Cam's method

We introduce the concepts of modeling adaptive protocols for quantum learning and distinguishing tasks with learning trees (Aharonov et al., 2022; Chen et al., 2022b, 2023a; Chen and Gong, 2025; Chen et al., 2024b). Here, we consider an arbitrary protocol using c-copy joint measurements and N copies, which is described by an element in $\mathcal{M}_{c,N}$. We split the protocol into N/c iterations, select c copies of the unknown state ρ at each node, perform a c-copy POVM in (a subset of) $\mathcal{M}_c$, and step to the next iteration corresponding to the outcome. We describe such a procedure with adaptivity using the learning tree representation adapted from Ref. (Chen et al., 2024b):

Definition 6 (Tree representation for protocols in $\mathcal{M}_{c,N}$ (Chen et al., 2024b)) *Given an unknown n-qubit quantum state ρ, a protocol using c-copy joint measurements and N copies of ρ in $\mathcal{M}_{c,N}$ can be represented as a rooted tree $\mathcal{T}$ of depth $T = N/c$ with each node on the tree recording the measurement outcome history of the algorithm. It has the following properties:*

1. *We assign a probability $p^\rho(u)$ to each node u on the tree $\mathcal{T}$. The probability assigned to the root r is $p^\rho(r) = 1$.*
2. *At each non-leaf node u, we measure a fresh batch $\rho^{\otimes c}$ containing c copies of ρ using a joint measurement $M_u = \{F_s^u\} \in \mathcal{M}_c$, resulting in a classical outcome s. Each child node v corresponding to the classical outcome s of the node u is connected through the edge $e_{u,s}$.*
3. *If a node v is the child of a node u through the edge $e_{u,s}$, the probability assigned to this edge is*

$$p^\rho(v) = p^\rho(u)\cdot \mathrm{tr}(F_s^u\rho^{\otimes c}). \tag{B.5}$$

4. *Each root-to-leaf path is of length $T = N/c$. At a leaf node ℓ, $p^{\rho}(\ell)$ denotes the probability of the classical memory reaching ℓ at the end of the protocol. We also denote the set of leaves of $\mathcal{T}$ by leaf($\mathcal{T}$).*

At the end of the protocol, the classical post-processing maps each leaf node to a desired output of the protocol.

Throughout this paper, we will use the learning tree representation defined in Theorem 6 as a tool to prove lower bounds for (oblivious) distinguishing tasks in Problem 3 and Problem 3'. In particular, we are interested in the following distinguishing task. Given access to copies of a d-dimensional unknown state ρ, the goal is to distinguish between the following two cases:

- (Null hypothesis) ρ is a state ρ_0; or
- (Alternative hypothesis) ρ is a parameterized state $\rho_{\theta,\varphi}$ randomly sampled from a probability distribution π.

Note that here we would like our distinguishing algorithm to apply to any choice of $\rho_0 \in \mathcal{S}^\circ$ and probability distribution π. That means for a fixed measurement protocol in $\mathcal{M}_{c,N}$, ρ_0, $\{Q_a\}_{a\in A}$, $\{T_b\}_{b\in B}$ and π can be chosen *adversarially* to reduce the successful rate.

There is a well-established framework for proving lower bounds of this distinguishing task, consisting of Le Cam's two-point method (Yu, 1997), one-sided likelihood ratio (Chen et al., 2022b), and the martingale technique (Chen et al., 2023a). We recap the necessary concepts here.

Definition 7 (Likelihood ratio) *Consider a protocol described by a tree representation $\mathcal{T}$ for the distinguishing task between the null and alternative hypotheses. For any leaf node $\ell \in$ leaf($\mathcal{T}$), we define the likelihood ratio to be*

$$L(\ell) := \frac{\mathbb{E}_{(\theta,\varphi)\sim\pi}[p^{\rho_{\theta,\varphi}}(\ell)]}{p^{\rho_0}(\ell)}. \tag{B.6}$$

We can also define the likelihood ratio for each edge $e_{u,s}$ and each particular choice of $\rho_{\theta,\varphi}$ as:

$$L_{\theta,\varphi}(\ell) := \frac{p^{\rho_{\theta,\varphi}}(\ell)}{p^{\rho_0}(\ell)}, \qquad L_{\theta,\varphi}(u,s) := \frac{p^{\rho_{\theta,\varphi}}(s|u)}{p^{\rho_0}(s|u)}. \tag{B.7}$$

We summarize the toolbox for showing lower bounds under this learning tree representation.

Lemma 8 (Toolbox of showing lower bounds) *Suppose $\mathcal{T}$ is a learning tree with depth $t = N/c$ that solves the distinguishing problem with probability p_{suc}.*

1. *(Le Cam's two-point method (Yu, 1997))*

$$p_{suc} \le d_{TV}(\mathbb{E}_{(\theta,\varphi)\sim\pi}[p^{\rho_{\theta,\varphi}}], p^{\rho_0}) = \frac{1}{2}\sum_{\ell\in leaf(\mathcal{T})} \left|\mathbb{E}_{(\theta,\varphi)\sim\pi}[p^{\rho_{\theta,\varphi}}(\ell)] - p^{\rho_0}(\ell)\right|. \tag{B.8}$$

2. *(One-sided likelihood ratio (Chen et al., 2022b)) For any $\beta > 0$, we have*

$$\begin{aligned} d_{TV}(\mathbb{E}_{(\theta,\varphi)\sim\pi}[p^{\rho_{\theta,\varphi}}], p^{\rho_0}) &\le \Pr_{\ell\sim p^{\rho_0},(\theta,\varphi)\sim\pi}[L_{\theta,\varphi}(\ell) \le \beta] + 1 - \beta, \\ d_{TV}(\mathbb{E}_{(\theta,\varphi)\sim\pi}[p^{\rho_{\theta,\varphi}}], p^{\rho_0}) &\le \Pr_{\ell\sim p^{\rho_0}}[L(\ell) \le \beta] + 1 - \beta. \end{aligned} \tag{B.9}$$

3. *(Martingale technique (Chen et al., 2023a, 2024b)) Suppose there is a $\delta > 0$ such that for every node u and outcome s we have*

$$\mathbb{E}_{(\theta,\varphi)\sim\pi}\mathbb{E}_{s\sim p^{\rho_0}(s|u)}\left[(L_{\theta,\varphi}(u,s)-1)^2\right]\leq\delta. \tag{B.10}$$

We then have

$$\Pr_{\ell\sim p^{\rho_0},(\theta,\varphi)\sim\pi}[L_{\theta,\varphi}(\ell)\leq 0.9]\leq 0.1+O(\delta t). \tag{B.11}$$

The proof of Lemma 8 can be referred to Refs. (Chen et al., 2022b, 2023a, 2024b). The first bound is a tree-based one, as we upper bound the total variation distance between the probability distributions of reaching each leaf under the two cases. The second bound is a path-based one, as we prove that the likelihood ratio is not too small for most of the paths from the root to leaves. The third bound is an edge-based one, as we focus on each edge and show that the likelihood ratio over the edge concentrates around 1.

Note that in Problem 3 and Problem 3', $\rho_{\theta,\varphi}$ is specified to be

$$\rho_{\theta,\varphi}=\rho_0+\frac{1}{d}\sum_{a\in A}\theta_a Q_a+\frac{1}{d}\sum_{b\in B}\varphi_b T_b, \tag{B.12}$$

where ρ_0 is the density matrix in the null hypothesis, $\{Q_a,T_b\}_{a\in A,b\in B}$ are dual observable bases, and $\theta\in\mathbb{R}^m$ and $\varphi\in\mathbb{R}^{d^2-m-1}$ can be arbitrary vectors such that $\rho_{\theta,\varphi}$ is well-defined and $\|\theta\|_p=3\varepsilon$. We denote the set of all such (θ,φ) as $\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$, i.e.

$$\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)=\{(\theta,\varphi)|\rho_{\theta,\varphi}\succeq 0,\|\theta\|_p=3\varepsilon,\theta\in\mathbb{R}^m,\varphi\in\mathbb{R}^{d^2-m-1}\}. \tag{B.13}$$

where $Q:=(Q_1\ Q_2\ \cdots\ Q_m)$ and $T:=(T_{m+1}\ T_{m+2}\ \cdots\ T_{d^2-1})$ represent the choice of the dual basis and the superscript ^{Q,T} highlights the dependence of (θ,φ) on the choice of basis Q and T. We also denote the set of all probability distributions over such (θ,φ) as $\mathcal{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$. Furthermore, we denote the set of all Q,T satisfying Eq. (2.1) as

$$\mathscr{B}(O)=\{(Q,T)|\mathrm{tr}(O_iQ_a)=d\delta_{ia},\forall i,a\in A,\mathrm{tr}(O_iT_b)=0,\forall i\in A,b\in B\}, \tag{B.14}$$

where $O:=(O_1\ O_2\ \cdots\ O_m)$.

Assume there is a protocol in $\mathcal{M}_{c,N}$ that solves this distinguishing problem in $T=N/c$ rounds. The algorithm can be represented by a learning tree $\mathcal{T}$ of depth T in Theorem 6. Let u be an internal node in the learning tree, and $M_u=\{F^u_s\}\in\mathcal{M}_c$ be the c-copy POVM used in the node. Then the probability of observing the outcome s given underlying state ρ is $p^\rho(s|u)=\mathrm{tr}(F^u_s\rho^{\otimes c})$. The likelihood ratio is thus $L_{\theta,\varphi}(u,s)=\mathrm{tr}(F^u_s\rho^{\otimes c}_{\theta,\varphi})/\mathrm{tr}(F^u_s\rho^{\otimes c}_0)$, and

$$\begin{aligned}\mathbb{E}_{(\theta,\varphi)\sim\pi}\mathbb{E}_{s\sim p^{\rho_0}(s|u)}\left[(L_{\theta,\varphi}(u,s)-1)^2\right]&=\mathbb{E}_{(\theta,\varphi)\sim\pi}\mathbb{E}_{s\sim p^{\rho_0}(s|u)}\left[\left(\frac{\mathrm{tr}(F^u_s\rho^{\otimes c}_{\theta,\varphi})}{\mathrm{tr}(F^u_s\rho^{\otimes c}_0)}-1\right)^2\right]\\&=\mathbb{E}_{(\theta,\varphi)\sim\pi}\chi^2_{M_u}\left(\rho^{\otimes c}_{\theta,\varphi}\|\rho^{\otimes c}_0\right)\end{aligned} \tag{B.15}$$

where $\chi^2_M(\rho\|\rho')$ denotes the χ^2-distance between the probability distribution over all measurement outcomes using the POVM M over ρ and ρ'. In particular, we can choose (Q,T), ρ_0 and π that (almost) minimize the last line so that the last line

$$\mathbb{E}_{(\theta,\varphi)\sim\pi}\chi^2_M\left(\rho_{\theta,\varphi}^{\otimes c}\|\rho_0^{\otimes c}\right) \leq 2 \inf_{\substack{\rho_0\in\mathcal{S}^\circ\\(Q,T)\in\mathscr{B}(O)}} \min_{\pi\in\mathcal{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \mathbb{E}_{(\theta,\varphi)\sim\pi}\chi^2_M\left(\rho_{\theta,\varphi}^{\otimes c}\|\rho_0^{\otimes c}\right). \tag{B.16}$$

Let

$$\delta_{\mathcal{M}_c}(O) := \sup_{M\in\mathcal{M}_c} \inf_{\substack{\rho_0\in\mathcal{S}^\circ\\(Q,T)\in\mathscr{B}(O)}} \min_{\pi\in\mathcal{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \mathbb{E}_{(\theta,\varphi)\sim\pi}\chi^2_M\left(\rho_{\theta,\varphi}^{\otimes c}\|\rho_0^{\otimes c}\right). \tag{B.17}$$

We note that there may be some $\rho_0\in\mathcal{S}^\circ$ such that $\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$ and $\mathcal{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$ are empty sets and the distinguishing task is not well defined. In this case, we let $\min_{\pi\in\mathcal{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} = \infty$. By Lemma 8, the sample complexity of this distinguishing task is lower bounded by $\Omega(c/\delta_{\mathcal{M}_c})$.

B.3. Quantum metrology

Here, we introduce basic concepts in quantum metrology (i.e. quantum estimation theory), including locally unbiased estimators, Cramér–Rao bound, and Fisher information matrix.

Consider a d-dimensional quantum state ρ_θ in Hilbert space $\mathcal{H}$, where $\theta=(\theta_1,\theta_2,\ldots,\theta_m)\in\Theta\subseteq\mathbb{R}^m$. Given a POVM $M=\{M_x\}$, an estimator $\hat{\theta}(x)$, a function that maps the measurement outcome to Θ is called an *unbiased* estimator if

$$\theta=\sum_x \hat{\theta}(x)\mathrm{tr}(\rho_\theta M_x) \tag{B.18}$$

for all $\theta\in\Theta$. A *locally unbiased* estimator describes an estimation that is unbiased in the vicinity of one specific value of θ, say θ_0, satisfying

$$\theta_0=\sum_x \hat{\theta}(x)\mathrm{tr}(\rho_\theta M_x)\bigg|_{\theta=\theta_0}, \quad \delta_{ij}=\frac{\partial}{\partial\theta_i}\sum_x \hat{\theta}_j(x)\mathrm{tr}(\rho_\theta M_x)\bigg|_{\theta=\theta_0}. \tag{B.19}$$

Unbiased estimators are locally unbiased, though the converse need not hold. In quantum metrology, the figure of merit is usually taken to be the mean square error matrix (MSEM), defined by

$$V(\rho_\theta,M,\hat{\theta})_{ij}=\sum_x(\hat{\theta}(x)_i-\theta_i)(\hat{\theta}(x)_j-\theta_j)\mathrm{tr}(\rho_\theta M_x). \tag{B.20}$$

We will also call $V(\rho_\theta,M,\hat{\theta})_{ii}$ the mean square error (MSE) for the estimator $\hat{\theta}_i$. For the estimator $\hat{\theta}_\alpha=\sum_i\alpha_i\hat{\theta}_i$, the corresponding MSE is given by

$$\sum_x(\hat{\theta}(x)_\alpha-\theta_\alpha)^2\mathrm{tr}(\rho_\theta M_x)=\alpha^\top V(\rho_\theta,M,\hat{\theta})\alpha. \tag{B.21}$$

Specifically, for locally unbiased estimators, it corresponds to the covariance matrix. The MSEM of any (locally) unbiased estimator (at θ) is bounded below by the inverse of the *Fisher information matrix* (FIM),

$$I(\rho_\theta,M)_{ij}:=\sum_{x,p_\theta(x):=\mathrm{tr}(\rho_\theta M_x)\neq0}\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta_i}\frac{\partial p_\theta(x)}{\partial\theta_j}, \tag{B.22}$$

through the Cramér–Rao (CR) bound,

$$V(\rho_\theta, M, \hat{\theta}) \succeq I(\rho_\theta, M)^{-1}, \tag{B.23}$$

where $V \succeq W$ means $V - W$ is positive semidefinite. Given N copies of quantum state ρ_θ, using the fact that the FIM is additive, we have

$$V(\rho_\theta, M^{\otimes N}, \hat{\theta}^{(N)}) \succeq I(\rho_\theta^{\otimes N}, M^{\otimes N})^{-1} = \frac{1}{N} I(\rho_\theta, M)^{-1}, \tag{B.24}$$

for any locally unbiased estimator $\hat{\theta}^{(N)}$.

The FIM is closely related to the second-order derivative of χ^2-distance. Consider the probability distribution $p_\theta(x)$. Assume $p_\theta(x) > 0$ for all x. Then,

$$\begin{aligned}\chi^2(p_{\theta+\mathrm{d}\theta} \| p_\theta) &= \sum_x p_\theta(x) \left(\frac{p_{\theta+\mathrm{d}\theta}(x)}{p_\theta(x)} - 1 \right)^2 = \sum_x \frac{(p_{\theta+\mathrm{d}\theta}(x) - p_\theta(x))^2}{p_\theta(x)} \\ &= \sum_x \frac{(\sum_i \partial_i p_\theta(x) \mathrm{d}\theta_i)^2}{p_\theta(x)} = \sum_{x,i,j} \frac{\partial_i p_\theta(x) \partial_j p_\theta(x) \mathrm{d}\theta_i \mathrm{d}\theta_j}{p_\theta(x)} = \sum_{i,j} I(M)_{ij} \mathrm{d}\theta_i \mathrm{d}\theta_j.\end{aligned} \tag{B.25}$$

In particular, when p_θ is linear in θ, the above calculation is exact for finite $\mathrm{d}\theta$, a situation that we will encounter later.

Consider the asymptotic situation where we have N copies of quantum state ρ_θ and $N \to \infty$. Under certain regularity conditions, the CR bound is saturable by maximum likelihood estimators (Van der Vaart, 2000) is the sense that

$$\sqrt{N}(\hat{\theta}^{(N)}_{\text{MLE}} - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, I(\rho_\theta, M)^{-1}), \tag{B.26}$$

i.e., $\sqrt{N}(\hat{\theta}^{(N)}_{\text{MLE}} - \theta)$ converges in distribution to a normal distribution centered around 0 with variance equal to $I(\rho_\theta, M)^{-1}$ as $N \to \infty$. When N is finite, it is unclear whether the CR bound is always saturable. However, when we focus on local estimation at a specific point θ_0 (whose value is known in prior), the following locally unbiased estimator automatically achieves the CR bound, i.e.,

$$V(\rho_\theta, M, \hat{\theta}^{\text{opt}})|_{\theta=\theta_0} = I(\rho_\theta, M)^{-1}|_{\theta=\theta_0}. \tag{B.27}$$

where for a measurement outcome y,

$$\hat{\theta}^{\text{opt}}(y; \theta_0) := (\theta_0)_i + \sum_x (\gamma_{i,x}|_{\theta=\theta_0}) \delta_{xy}, \quad \gamma_{i,x} = \sum_j (I(\rho_\theta, M)^{-1})_{ij} \frac{\partial_j p_\theta(x)}{p_\theta(x)}. \tag{B.28}$$

We can easily verify that this estimator achieves the CR bound:

$$\begin{aligned}V(\rho_\theta, M, \hat{\theta}^{\text{opt}})_{ij} &= \sum_x (\hat{\theta}^{\text{opt}}(x;\theta)_i - \theta_i)(\hat{\theta}^{\text{opt}}(x;\theta)_j - \theta_j) \mathrm{tr}(\rho_\theta M_x) \\ &= \sum_{x,x_1,x_2} \gamma_{i,x_1} \delta_{x_1,x} \gamma_{j,x_2} \delta_{x_2,x} p_\theta(x) = \sum_x \gamma_{i,x} \gamma_{j,x} p_\theta(x) \\ &= \sum_{i'j'} (I(\rho_\theta, M)^{-1})_{ii'} \frac{\partial_{i'} p_\theta(x)}{p_\theta(x)} (I(\rho_\theta, M)^{-1})_{jj'} \frac{\partial_{j'} p_\theta(x)}{p_\theta(x)} p_\theta(x) \\ &= \sum_{i'j'} (I(M)^{-1})_{ii'} I(M)_{i'j'} (I(M)^{-1})_{jj'} = (I(\rho_\theta, M)^{-1})_{ij}.\end{aligned} \tag{B.29}$$

In practice, however, the optimal locally unbiased estimator Eq. (B.28) cannot be directly applied as the local point θ_0 is unknown in prior.

B.4. Tail bounds

We will need the following Chebyshev's inequality and Hoeffding's inequality.

Lemma 9 (Chebyshev's inequality, see e.g., Corollary 1.6.3 in (Vershynin, 2018)) *Let X be a random variable with finite non-zero variance σ^2 and finite expected value μ. Then for any real number $k > 0$, we have*

$$\Pr[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2}. \tag{B.30}$$

Lemma 10 (Hoeffding's inequality (Hoeffding, 1963), e.g., Theorem 2.2.6 in (Vershynin, 2018)) *Let $X_1, \cdots, X_M$ be M independent random variables such that $X_i \in [a_i, b_i]$ for every i. For any $t > 0$, we have*

$$\Pr\left[\left|\sum_{i=1}^{M}(X_i - \mathbb{E}[X_i])\right| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{M}(b_i-a_i)^2}\right). \tag{B.31}$$

As a corollary, let $X_1, \cdots, X_M$ be i.i.d. random variables in $[0,1]$ and $\mathbb{E}[X_i] = \mu$ for every i. Then

$$\Pr\left[\sum_{i=1}^{M} X_i \leq (1-\delta)M\mu\right] \leq e^{-2M\mu^2\delta^2}. \tag{B.32}$$

Appendix C. Lower bounds for many-versus-one distinguishing

In this section, we prove a lower bound for the (oblivious) many-versus-one distinguishing problem in Problem 3 (Problem 3'). We recap the distinguishing task we consider here for concreteness. Given access to copies of a d-dimensional unknown state ρ, the goal is to distinguish between the following two cases:

- (Null hypothesis) ρ is a state ρ_0; or
- (Alternative hypothesis) ρ is a parameterized state $\rho_{\theta,\varphi}$ randomly sampled from a probability distribution $\pi \in \mathcal{D}_{3\varepsilon,p}^{Q,T}(\rho_0)$. Here, $\rho_{\theta,\varphi}$ is defined to be

  $$\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b, \tag{C.1}$$

  where ρ_0 is the density matrix in the null hypothesis, $\{Q_a, T_b\}_{a\in A, b\in B}$ are dual observable bases, and $\theta \in \mathbb{R}^m$ and $\varphi \in \mathbb{R}^{d^2-m-1}$ can be arbitrary vectors such that $\rho_{\theta,\varphi}$ is well-defined and $\|\theta\|_p = 3\varepsilon$.

If $B = \varnothing$ and $\{Q_a\}_{a\in A}$ forms a complete basis of the Hilbert space, we called $\{Q_a\}_{a\in A}$ a set of complete observables. Recall from the learning tree framework, we showed in Eq. (B.17) that any protocol described by $\mathcal{M}_{c,N}$ in Theorem 6 requires $\Omega(c/\delta_{\mathcal{M}_c})$ samples with

$$\delta_{\mathcal{M}_c}(O) = \sup_{M\in\mathcal{M}_c} \inf_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \min_{\pi\in\mathcal{D}_{3\varepsilon,p}^{Q,T}(\rho_0)} \mathbb{E}_{(\theta,\varphi)\sim\pi}\chi_M^2\left(\rho_{\theta,\varphi}^{\otimes c} \| \rho_0^{\otimes c}\right). \tag{C.2}$$

C.1. Single-copy measurements: Exact lower bounds

We first assume that we can only use single-copy measurements ($c = 1$).

C.1.1. COMPLETE OBSERVABLES

We start with the case when the observables form a complete basis. The task reduces to distinguishing between the following two cases:

- (Null hypothesis) ρ is a state ρ_0; or
- (Alternative hypothesis) ρ is a parameterized state ρ_θ randomly sampled from a probability distribution $\pi \in \mathcal{D}^{Q}_{3\varepsilon,p}(\rho_0)$. Here, ρ_θ is defined to be

$$\rho_\theta = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a, \tag{C.3}$$

where ρ_0 is the density matrix in the null hypothesis, $\{Q_a\}_{a\in A}$ forms a complete observable basis, and $\theta \in \mathbb{R}^m$ can be arbitrary vectors such that ρ_θ is well-defined and $\left\|\theta\right\|_p = 3\varepsilon$.

Theorem 11 (Lower bound for Problem 3(3') with complete observables, $c = 1$, and p-norm error) *Using the adaptive measurement strategy with single-copy measurements, the sample complexity required to solve the many-versus-one distinguishing tasks above with any ρ_θ well-defined and $\theta \in \mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)$ is*

$$N = \Omega\left(\inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}^\circ}\max_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)}\left(\theta^\top I(\rho_0, M)\theta\right)^{-1}\right) \tag{C.4}$$

for any $\varepsilon > 0$, $p \in [1,\infty]$, where $I(\rho_0, M) := I(\rho_\theta, M)|_{\theta=\mathbf{0}}$ and $\rho_\theta = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a$.

Proof The sample complexity lower bound from Eq. (B.17) is then given by $\Omega(1/\delta_{\mathcal{M}}(O))$ where

$$\delta_{\mathcal{M}}(O) = \sup_{M\in\mathcal{M}}\inf_{\rho_0\in\mathcal{S}^\circ}\min_{\pi\in\mathcal{D}^{Q}_{3\varepsilon,p}(\rho_0)}\mathbb{E}_{\theta\sim\pi}\chi^2_M\left(\rho_\theta\|\rho_0\right). \tag{C.5}$$

Note that $\{Q_a\}_{a\in A}$ here can be uniquely defined using $\{O_i\}_{i=1}^m$. Also note that the optimal π distribution in this expression would trivially be the delta function at the value of θ that minimizes the χ^2-distance. Thus, we have

$$\delta_{\mathcal{M}}(O) = \sup_{M\in\mathcal{M}}\inf_{\rho_0\in\mathcal{S}^\circ}\min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)}\chi^2_M\left(\rho_\theta\|\rho_0\right). \tag{C.6}$$

Given a fixed measurement $M = \{M_s\} \in \mathcal{M}$ and denote $p_s = \mathrm{tr}(\rho_0 M_s)$, we compute the minimization part above as:

$$\min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)}\chi^2_M(\rho_\theta\|\rho_0) = \min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)}\sum_s p_s\left(\frac{\mathrm{tr}(M_s\rho_\theta)}{p_s} - 1\right)^2. \tag{C.7}$$

We can extend $\mathrm{tr}(M_s\rho_\theta)$ as $\mathrm{tr}(M_s\rho_\theta) = p_s + \sum_i \theta_i \partial_i p_s$. We thus have

$$\begin{aligned}\min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \chi^2_M(\rho_\theta\|\rho_0) &= \min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \sum_s p_s \left(\frac{\mathrm{tr}(M_s\rho_\theta)}{p_s} - 1\right)^2 \\ &= \min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \sum_s \frac{(\sum_i \theta_i\partial_i p_s)^2}{p_s} \\ &= \min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \sum_{s,i,j} \frac{\theta_i\theta_j\partial_i p_s \partial_j p_s}{p_s} \\ &= \min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \theta^\top I(\rho_0, M)\theta.\end{aligned} \tag{C.8}$$

Therefore, we have

$$\delta_{\mathcal{M}}(O) = \sup_{M\in\mathcal{M}} \inf_{\rho_0\in\mathcal{S}^\circ} \min_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \theta^\top I(\rho_0, M)\theta. \tag{C.9}$$

Taking the inverse, the sample complexity lower bound is then given by

$$N = \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \left(\theta^\top I(\rho_0, M)\theta\right)^{-1}\right) \tag{C.10}$$

as claimed. ■

C.1.2. A GENERAL SET OF OBSERVABLES

We then consider a general dual observable basis $Q = \{Q_a\}_{a\in A}$ and $T = \{T_b\}_{b\in B}$ with single-copy measurement protocols ($c = 1$). We show the following sample complexity lower bound:

Theorem 12 (Lower bound for Problem 3(3’) at $c = 1$ and p-norm error) *Using the adaptive measurement strategy with single-copy measurements, the sample complexity required to solve the many-versus-one distinguishing tasks above with any $\rho_{\theta,\varphi}$ well-defined and $(\theta,\varphi) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$ is*

$$N = \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \max_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \left((\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi)\right)^{-1}\right) \tag{C.11}$$

for any $\varepsilon > 0$, $p \in [1,\infty]$, where $I(\rho_0, M) = I(\rho_{\theta,\varphi}, M)|_{\theta=\varphi=\mathbf{0}}$ and $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b$.

Proof The sample complexity lower bound from Eq. (B.17) is then given by $\Omega(1/\delta_{\mathcal{M}}(O))$ where

$$\begin{aligned}\delta_{\mathcal{M}}(O) &= \sup_{M\in\mathcal{M}} \inf_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \min_{\pi\in\mathcal{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \mathbb{E}_{(\theta,\varphi)\sim\pi}\chi^2_M\left(\rho_{\theta,\varphi}\|\rho_0\right) \\ &= \sup_{M\in\mathcal{M}} \inf_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \chi^2_M\left(\rho_{\theta,\varphi}\|\rho_0\right)\end{aligned} \tag{C.12}$$

Given a fixed measurement $M = \{M_s\} \in \mathcal{M}$ and denote $p_s = \operatorname{tr}(\rho_0 M_s)$, we compute the minimization part above as:

$$\min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \chi^2_M(\rho_{\theta,\varphi}\|\rho_0) = \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \sum_s p_s \left(\frac{\operatorname{tr}(M_s\rho_{\theta,\varphi})}{p_s} - 1\right)^2. \tag{C.13}$$

We can extend $\operatorname{tr}(M_s\rho_{\theta,\varphi})$ as $\operatorname{tr}(M_s\rho_\theta) = p_s + \sum_i \theta_i\partial_i p_s + \sum_{i'} \varphi_{i'}\partial_{i'} p_s$. We thus have

$$\begin{aligned}
\min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \chi^2_M(\rho_\theta\|\rho_0) &= \min_{\theta\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \sum_s p_s \left(\frac{\operatorname{tr}(M_s\rho_\theta)}{p_s} - 1\right)^2 \\
&= \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \sum_s \frac{\left(\sum_i \theta_i\partial_i p_s + \sum_{i'} \varphi_{i'}\partial_{i'} p_s\right)^2}{p_s} \\
&= \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi).
\end{aligned} \tag{C.14}$$

Therefore, we have

$$\delta_{\mathcal{M}}(O) = \sup_{M\in\mathcal{M}} \inf_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi). \tag{C.15}$$

Taking the inverse, the sample complexity lower bound is then given by

$$N = \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \max_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \left((\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi)\right)^{-1} \right) \tag{C.16}$$

as claimed. ■

C.2. Duality between many-versus-one distinguishing and parameter estimation

Before going further to explore the lower bounds in Theorem 11 and Theorem 12, we first reveal a key relationship between the lower bounds of these many-versus-one distinguishing tasks and parameter estimation tasks.

In the complete observables case, from Theorem 11, we have shown that the sample lower bound for distinguishing between ρ_0 versus ρ_θ is given by

$$N = \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\theta\in\mathscr{D}^{Q}_{3\varepsilon,p}(\rho_0)} \left(\theta^\top I(\rho_0, M)\theta\right)^{-1} \right) \tag{C.17}$$

for any $\varepsilon > 0$, $p \in [1, \infty]$, and ρ_0. However, this sample complexity lower bound in the form

$$\max_\theta \left(\theta^\top I(\rho_0, M)\theta\right)^{-1} \tag{C.18}$$

is different from the usual lower (upper) bound in parameter estimation, which is in the form of

$$\max_\alpha \alpha^\top I(\rho_0, M)^{-1}\alpha. \tag{C.19}$$

The latter represents a lower bound on the MSE that can be naturally derived from the CR bound (Eq. (B.23)) using Eq. (B.21) in parameter estimation.

A similar situation appears for the case of general observables. The sample complexity lower bound for distinguishing between ρ_0 versus $\rho_{\theta,\varphi}$ from Theorem 12 is given by

$$N = \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \max_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \left((\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)\right)^{-1}\right) \tag{C.20}$$

for any $\varepsilon > 0$, $p \in [1,\infty]$, and ρ_0. This is again in the form of

$$\max_{\theta,\varphi} \left((\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)\right)^{-1}, \tag{C.21}$$

which is different from the lower (upper) bound in parameter estimation in the form of

$$\max_{\alpha} (\alpha,0)^\top I(\rho_0,M)^{-1}(\alpha,0). \tag{C.22}$$

Surprisingly, we show the following duality relationship between many-versus-one distinguishing and parameter estimation using the following mathematical fact, by temporarily extending the domains of each optimization to the entire space with bounded p- and q-norms.

Lemma 13 (Duality between many-versus-one distinguishing and parameter estimation) *Let the FIM $I(\rho_0,M)$ be a $(|A|+|B|)\times(|A|+|B|)$ positive semi-definite matrix. We use (θ,φ) to represent a joint vector whose first $|A|$ columns are θ and last $|B|$ columns are φ. We have*

$$\begin{aligned}
&\max_{\|\theta\|_p\geq 1,\theta\in\mathbb{R}^{|A|},\varphi\in\mathbb{R}^{|B|}} \left((\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)\right)^{-1} \\
=&\max_{\|\alpha\|_q\leq 1,\alpha\in\mathbb{R}^{|A|}} (\alpha,0)^\top I(\rho_0,M)^{-1}(\alpha,0) \\
=&\max_{\|\alpha\|_q\leq 1,\alpha\in\mathbb{R}^{|A|}} \alpha^\top (I(\rho_0,M)^{-1})_{AA}\alpha
\end{aligned} \tag{C.23}$$

for $1/p+1/q=1$, where the subscript $_{AA}$ represents the upper-left block of the FIM restricted to $a\in A$.

Proof We only need to show

$$\begin{aligned}
&\min_{\|\theta\|_p\geq 1,\theta\in\mathbb{R}^{|A|},\varphi\in\mathbb{R}^{|B|}} (\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi) \\
=&\frac{1}{\max_{\|\alpha\|_q\leq 1,\alpha\in\mathbb{R}^{|A|}}(\alpha,0)^\top I(\rho_0,M)^{-1}(\alpha,0)}.
\end{aligned} \tag{C.24}$$

Assume $(\theta^\star,\varphi^\star)$ minimizes $(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)$, $\|\theta^\star\|_p = 1$. Then there exists $\alpha^\diamond$ such that $|\alpha^\diamond\cdot\theta^\star| = 1$ and $\|\alpha^\diamond\|_q \leq 1$. The Cauchy-Schwarz inequality implies

$$\begin{aligned}
(\alpha^\diamond,0)^\top I(\rho_0,M)^{-1}(\alpha^\diamond,0) \geq&\frac{1}{(\theta^\star,\varphi^\star)^\top I(\rho_0,M)(\theta^\star,\varphi^\star)} \\
=&\frac{1}{\min_{\|\theta\|_p\geq 1,\theta\in\mathbb{R}^{|A|},\varphi\in\mathbb{R}^{|B|}}(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)},
\end{aligned} \tag{C.25}$$

which means the left-hand side of Eq. (C.24) is no smaller than the right-hand side.

On the other hand, assume $\alpha^\star$ maximizes $(\alpha,0)^\top I(\rho_0,M)^{-1}(\alpha,0)$, which must satisfy $\|\alpha^\star\|_q = 1$. Let $(\theta^\diamond,\varphi^\diamond) = \frac{I(\rho_0,M)^{-1}(\alpha^\star,0)}{\|(I(\rho_0,M)^{-1}(\alpha^\star,0))_A\|_p}$, we have

$$
\begin{aligned}
(\theta^\diamond,\varphi^\diamond)^\top I(\rho_0,M)(\theta^\diamond,\varphi^\diamond) &= \frac{(\alpha^\star,0)^\top I(\rho_0,M)^{-1}(\alpha^\star,0)}{\|(I(\rho_0,M)^{-1}(\alpha^\star,0))_A\|_p^2} \\
&= \frac{(\alpha^\star,0)^\top I(\rho_0,M)^{-1}(\alpha^\star,0)}{\|\alpha^\star\|_q^2\|(I(\rho_0,M)^{-1}(\alpha^\star,0))_A\|_p^2} \\
&\leq \frac{(\alpha^\star,0)^\top I(\rho_0,M)^{-1}(\alpha^\star,0)}{((\alpha^\star,0)^\top I(\rho_0,M)^{-1}(\alpha^\star,0))^2} \\
&= \frac{1}{(\alpha^\star,0)^\top I(\rho_0,M)^{-1}(\alpha^\star,0)},
\end{aligned}
\tag{C.26}
$$

which means the left-hand side of Eq. (C.24) is no larger than the right-hand side. Here we use $(\cdot)_A$ to denote the first $|A|$ columns of a vector, and we use the generalized Cauchy-Schwarz: $\|\beta\|_p\|\alpha\|_q \geq |\beta\cdot\alpha|$. ■

C.3. Single-copy measurements: Threshold on ε to saturate the upper bound

Lemma 13 has shown that

$$
\begin{aligned}
&\min_{\|\theta\|_p\geq 1,\theta\in\mathbb{R}^{|A|},\varphi\in\mathbb{R}^{|B|}} (\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi) \\
=&\frac{1}{\max_{\|\alpha\|_q\leq 1,\alpha\in\mathbb{R}^{|A|}}(\alpha,0)^\top I(\rho_0,M)^{-1}(\alpha,0)}
\end{aligned}
\tag{C.27}
$$

for $1/p+1/q=1$. However, a more detailed observation on Theorem 11 and Theorem 12 indicates that the duality cannot directly work as the maximization over (θ,φ) is over $\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$ instead of simply the sphere $\|\theta\|_p = 3\varepsilon$.

Here, we propose a threshold η of ε such that the lower bounds in Theorem 11 and Theorem 12 can be written in the form of maximization over $\|\theta\|_p = 3\varepsilon$.

C.3.1. COMPLETE OBSERVABLES

We start with the case of complete observables in Theorem 11. We have shown that the lower bound is given by $\Omega(1/\delta_{\mathcal{M}}(O))$, where $\delta_{\mathcal{M}}(O)$ is defined by

$$
\delta_{\mathcal{M}}(O) = \sup_{M\in\mathcal{M}}\inf_{\rho_0\in\mathcal{S}^\circ}\min_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)}\theta^\top I(\rho_0,M)\theta. \tag{C.28}
$$

Our goal is to apply the duality result in Lemma 13 to the above lower bound. To this end, we prove the following result.

Theorem 14 (High-precision threshold with complete observables, $c=1$, and p-norm error) *Using the adaptive measurement strategy with single-copy measurements, the sample complexity* *required to solve the many-versus-one distinguishing task (Problem 3') between any well-defined ρ_θ ($\theta \in \mathscr{D}^Q_{3\varepsilon,p}(\rho_0)$) and ρ_0 is*

$$\begin{aligned} N &= \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\|\theta\|_p=1} \frac{1}{\varepsilon^2}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1}\right) \\ &= \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\|\alpha\|_q\le 1} \frac{1}{\varepsilon^2}\alpha^\top I(\rho_0,M)^{-1}\alpha\right) \end{aligned} \tag{C.29}$$

with any

$$\varepsilon \le \eta = \frac{1}{6\left\|(\|Q_1\|_\infty,\ldots,\|Q_{d^2-1}\|_\infty)\right\|_q}, \tag{C.30}$$

and for any $p\in[1,\infty]$, where $I(\rho_0,M) = I(\rho_\theta,M)|_{\theta=\mathbf{0}}$ and $\rho_\theta = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a$.

Proof Our hope is to show that there is some threshold η such that when $\varepsilon < \eta$, we have

$$\begin{aligned} \min_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)} \theta^\top I(\rho_0,M)\theta &= \min_{\|\theta\|_p=3\varepsilon} \theta^\top I(\rho_0,M)\theta \\ &= 9\varepsilon^2 \min_{\|\theta\|_p=1} \theta^\top I(\rho_0,M)\theta \\ &= \frac{9\varepsilon^2}{\max_{\|\alpha\|_q\le 1}\alpha^\top I(\rho_0,M)^{-1}\alpha} \end{aligned} \tag{C.31}$$

where the last step is due to the duality result in Lemma 13.

Unfortunately, there does not exist such a choice of η that applies to all ρ_0, because when ρ_0 is close to the boundary of the density matrix space, e.g., when ρ_0 is singular, $\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)$ cannot holds for all θ satisfying $\|\theta\|_p = 3\varepsilon$ even when ε is small. Luckily, we can restrict the range of ρ_0 without significantly increasing the value of

$$\delta_{\mathcal{M}}(O) = \sup_{M\in\mathcal{M}} \inf_{\rho_0\in\mathcal{S}^\circ} \min_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)} \chi^2_M(\rho_\theta\|\rho_0). \tag{C.32}$$

Defining

$$\mathcal{S}_{1/2} := \left\{\rho\middle|\rho = \frac{1}{2}\Big(\sigma + \frac{\mathbb{I}}{d}\Big),\ \text{for some density matrix } \sigma\right\}, \tag{C.33}$$

and using

$$\begin{aligned} &\theta^\top I\left(\frac{1}{2}\sigma + \frac{\mathbb{I}}{2d}, M\right)\theta = \chi^2_M\left(\frac{1}{2}\sigma + \frac{\mathbb{I}}{2d} + \frac{1}{d}\sum_a\theta_a Q_a\middle\|\frac{1}{2}\sigma + \frac{\mathbb{I}}{2d}\right) \\ &= \sum_s \frac{\mathrm{tr}(M_s\frac{1}{d}\sum_a\theta_a Q_a)^2}{\mathrm{tr}(M_s(\frac{1}{2}\sigma + \frac{\mathbb{I}}{2d}))} \le \sum_s \frac{\mathrm{tr}(M_s\frac{1}{d}\sum_a\theta_a Q_a)^2}{\mathrm{tr}(M_s(\frac{1}{2}\sigma + \frac{1}{2d}\sigma))} \\ &= \frac{2d}{d+1}\sum_s \frac{\mathrm{tr}(M_s\frac{1}{d}\sum_a\theta_a Q_a)^2}{\mathrm{tr}(M_s\sigma)} \\ &\le 2\sum_s \frac{\mathrm{tr}(M_s\frac{1}{d}\sum_a\theta_a Q_a)^2}{\mathrm{tr}(M_s\sigma)} = 2\chi^2_M(\sigma_\theta\|\sigma) = 2\theta^\top I(\sigma,M)\,\theta, \end{aligned} \tag{C.34}$$

where $\sigma_\theta = \sigma + \frac{1}{d}\sum_a \theta_a Q_a$, we have both

$$\begin{aligned} 2\inf_{\rho_0\in\mathcal{S}_{1/2}}\min_{\theta\in\mathscr{D}^Q_{3\varepsilon/2,p}(\rho_0)}\theta^\top I\left(\rho_0,M\right)\theta &\le \inf_{\rho_0\in\mathcal{S}^\circ}\min_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)}\theta^\top I\left(\rho_0,M\right)\theta \\ &\le \inf_{\rho_0\in\mathcal{S}_{1/2}}\min_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)}\theta^\top I\left(\rho_0,M\right)\theta, \end{aligned} \tag{C.35}$$

because $2\theta \in \mathscr{D}^Q_{3\varepsilon,p}(\rho_0)$ implies $\theta \in \mathscr{D}^Q_{3\varepsilon/2,p}(\rho_0/2 + \mathbb{I}/2d)$, and

$$\begin{aligned} \frac{1}{2}\inf_{\rho_0\in\mathcal{S}_{1/2}}\min_{\|\theta\|_p=3\varepsilon}\theta^\top I\left(\rho_0,M\right)\theta &\le \inf_{\rho_0\in\mathcal{S}^\circ}\min_{\|\theta\|_p=3\varepsilon}\theta^\top I\left(\rho_0,M\right)\theta \\ &\le \inf_{\rho_0\in\mathcal{S}_{1/2}}\min_{\|\theta\|_p=3\varepsilon}\theta^\top I\left(\rho_0,M\right)\theta. \end{aligned} \tag{C.36}$$

Now we try to find an η such that when $\varepsilon < \eta$, $\|\theta\|_p = 3\varepsilon$ implies $\theta \in \mathscr{D}^Q_{3\varepsilon,p}(\rho_0)$ for all $\rho_0 \in \mathcal{S}_{1/2}$. For example, we can choose η as claimed

$$\eta := \frac{1}{6\left\|\left(\|Q_1\|_\infty,\ldots,\|Q_{d^2-1}\|_\infty\right)\right\|_q}, \tag{C.37}$$

such that

$$\left\|\sum_a \theta_a Q_a\right\|_\infty \le \sum_a |\theta_a|\|Q_a\|_\infty \le \|\theta\|_p\left\|\left(\|Q_1\|_\infty,\ldots,\|Q_{d^2-1}\|_\infty\right)\right\|_q \le \frac{1}{2}. \tag{C.38}$$

This makes sure when $\rho_0 \in \mathcal{S}_{1/2}$, for any $\|\theta\|_p = 3\varepsilon$, ρ_θ is well-defined. Thus when $\varepsilon \le \eta$,

$$\inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}^\circ}\max_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1} \le \inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}^\circ}\max_{\|\theta\|_p=3\varepsilon}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1} \tag{C.39}$$

by definition, and

$$\begin{aligned} &\inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}^\circ}\max_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1} \\ \ge &\inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}_{1/2}}\max_{\theta\in\mathscr{D}^Q_{3\varepsilon,p}(\rho_0)}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1} \\ = &\inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}_{1/2}}\max_{\|\theta\|_p=3\varepsilon}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1} \\ \ge &\frac{1}{2}\inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}^\circ}\max_{\|\theta\|_p=3\varepsilon}\left(\theta^\top I(\rho_0,M)\theta\right)^{-1}. \end{aligned} \tag{C.40}$$

The theorem is then proved using Lemma 13. ■

C.3.2. A GENERAL SET OF OBSERVABLES

We also compute a threshold for the case of a general set of observables in the distinguishing task between ρ_0 and $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_a \theta_a Q_a + \frac{1}{d}\sum_b \varphi_b T_b$, where and $(\theta,\varphi) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$. To so do, we will explore different choices of dual basis $Q = \{Q_a\}_{a\in A}$ and $T = \{T_b\}_{b\in B}$ and pick a suitable one where a threshold can be easily obtain, as the dual basis is no longer uniquely defined as in the case of complete observables.

We first prove the following lemma.

Lemma 15 (Independence of the lower bound on the choice of dual basis) *The quantity*

$$(I(\rho_0, M)^{-1})_{AA} \tag{C.41}$$

is invariant under different choices of Q, T as long as Eq. (2.1) holds.

Proof Let both $(Q\ T) := (Q_1\ Q_2\ \cdots\ Q_m\ T_{m+1}\ T_{m+2}\ \cdots\ T_{d^2-1})$ and $(Q'\ T') := (Q'_1\ Q'_2\ \cdots\ Q'_m\ T'_{m+1}\ T'_{m+2}\ \cdots\ T'_{d^2-1})$ be valid choices of dual basis that satisfy Eq. (2.1), i.e.,

$$\operatorname{tr}(O_i Q_a) = d\delta_{ia}, \quad \operatorname{tr}(O_i T_b) = 0, \tag{C.42}$$

where we view $(Q\ T)$ (or $(Q'\ T')$) as a $d \times (d(d^2-1))$ block row vector where each block is a $d \times d$ dual observable. Then they must be related by the following:

$$\begin{pmatrix} Q' & T' \end{pmatrix} = \begin{pmatrix} Q + TC_1 & TC_2 \end{pmatrix} = \begin{pmatrix} Q & T \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ C_1 & C_2 \end{pmatrix}, \tag{C.43}$$

where we $C_1 \in \mathbb{R}^{|B|\times|A|} = \mathbb{R}^{d^2-1-m\times m}$ and $C_2 \in= \mathbb{R}^{|B|\times|B|} = \mathbb{R}^{d^2-1-m\times d^2-1-m}$ to represent linear transformations on the matrix blocks. C_1 can be arbitrary and C_2 needs to be invertible. Furthermore,

$$\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_a \theta_a Q_a + \frac{1}{d}\sum_b \varphi_b T_b = \rho_0 + \frac{1}{d}\sum_a \theta'_a Q'_a + \frac{1}{d}\sum_b \varphi'_b T'_b, \tag{C.44}$$

where

$$\begin{pmatrix} \theta' \\ \varphi' \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ C_1 & C_2 \end{pmatrix}^{-1} \begin{pmatrix} \theta \\ \varphi \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ -C_2^{-1}C_1 & C_2^{-1} \end{pmatrix} \begin{pmatrix} \theta \\ \varphi \end{pmatrix} = \begin{pmatrix} \theta \\ -C_2^{-1}C_1\theta + C_2^{-1}\varphi \end{pmatrix}. \tag{C.45}$$

As a result, the corresponding FIMs are related by the following:

$$I(\rho_0, M)' = \begin{pmatrix} \mathbb{I} & C_1^\top \\ 0 & C_2^\top \end{pmatrix} \begin{pmatrix} I(\rho_0, M)_{AA} & I(\rho_0, M)_{AB} \\ I(\rho_0, M)_{BA} & I(\rho_0, M)_{BB} \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ C_1 & C_2 \end{pmatrix}. \tag{C.46}$$

Since $(I(\rho_0, M)^{-1})_{AA}$ is equal to the inverse of the Schur complement of $I(\rho_0, M)$, i.e.

$$(I(\rho_0, M)^{-1})_{AA} = (I(\rho_0, M)_{AA} - I(\rho_0, M)_{AB} I(\rho_0, M)^{-1}_{BB} I(\rho_0, M)_{BA})^{-1}, \tag{C.47}$$

we only need to show the Schur complement $I(\rho_0, M)_{AA} - I(\rho_0, M)_{AB} I(\rho_0, M)_{BB}^{-1} I(\rho_0, M)_{BA}$ is invariant under the basis transformation. This can be seen from Eq. (C.46).

$$\begin{aligned} &I(\rho_0, M)'_{AA} - I(\rho_0, M)'_{AB} I(\rho_0, M)'_{BB}{}^{-1} I(\rho_0, M)'_{BA} \\ &= I(\rho_0, M)_{AA} + C_1^\top I(\rho_0, M)_{BA} + I(\rho_0, M)_{AB} C_1 + C_1^\top I(\rho_0, M)_{BB} C_1 \\ &\quad - (I(\rho_0, M)_{AB} + C_1^\top I(\rho_0, M)_{BB}) I(\rho_0, M)_{BB}^{-1} (I(\rho_0, M)_{BA} + I(\rho_0, M)_{BB} C_1) \\ &= I(\rho_0, M)_{AA} - I(\rho_0, M)_{AB} I(\rho_0, M)_{BB}{}^{-1} I(\rho_0, M)_{BA}. \end{aligned} \tag{C.48}$$

■

Now we define

$$\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m) := \inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \le 1} \alpha^\top (I(\rho_0, M))_{AA}^{-1} \alpha, \tag{C.49}$$

which is only a function of observables $\{O_i\}_{i=1}^m$ because the expression is independent of different choices of the dual basis. Below, we show that there is a unique choice of basis that allows us to derive the following threshold result.

Theorem 16 (High-precision threshold with general observables, $c = 1$, and p-norm error) *Using the adaptive measurement strategy with single-copy measurements, we consider the many-versus-one distinguishing task (Problem 3') between any well-defined $\rho_{\theta,\varphi}$ ($(\theta, \varphi) \in \mathscr{D}_{3\varepsilon,p}^{Q,T}(\rho_0)$) and ρ_0. The sample complexity required to solve it is*

$$\begin{aligned} N &= \Omega\left(\inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\theta\|_p = 1} \frac{1}{\varepsilon^2} \left((\theta, \varphi)^\top I(\rho_0, M) (\theta, \varphi) \right)^{-1} \right) \\ &= \Omega\left(\inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \le 1} \frac{1}{\varepsilon^2} (\alpha, 0)^\top I(\rho_0, M)^{-1} (\alpha, 0) \right) = \Omega\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{\varepsilon^2} \right), \end{aligned} \tag{C.50}$$

with any

$$\varepsilon \le \eta^{\mathrm{ob}} := \inf_{\rho_0 \in \mathcal{S}_{1/2}} \frac{1}{6 \left\| \left(\|Q'_1(\rho_0)\|_\infty, \ldots, \left\| Q'_{|A|}(\rho_0) \right\|_\infty \right) \right\|_q}, \tag{C.51}$$

and for any $p \in [1, \infty]$, where $I(\rho_0, M) := I(\rho_{\theta,\varphi}, M)|_{\theta=\varphi=\mathbf{0}}$ and $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d} \sum_{a \in A} \theta_a Q_a + \frac{1}{d} \sum_{b \in B} \varphi_b T_b$.

Here $Q'(\rho_0)$ is a special choice of Q as a function of ρ_0. It can be found by first picking arbitrary Q and T satisfying Eq. (2.1), and then picking $M^\star \in \mathcal{M}$ satisfying

$$\begin{aligned} &\inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}_{1/2}} \max_{(\theta,\varphi) \in \mathscr{D}_{3\varepsilon,p}^{Q',T}(\rho_0)} \left((\theta, \varphi)^\top I(\rho_0, M)' (\theta, \varphi) \right)^{-1} \ge \\ &\qquad \frac{1}{2} \sup_{\rho_0 \in \mathcal{S}_{1/2}} \max_{(\theta,\varphi) \in \mathscr{D}_{3\varepsilon,p}^{Q',T}(\rho_0)} \left((\theta, \varphi)^\top I(\rho_0, M^\star)' (\theta, \varphi) \right)^{-1}, \end{aligned} \tag{C.52}$$

where

$$I(\rho_0, M)' = \begin{pmatrix} \mathbb{I} & C_1^\top \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} I(\rho_0, M)_{AA} & I(\rho_0, M)_{AB} \\ I(\rho_0, M)_{BA} & I(\rho_0, M)_{BB} \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ C_1 & \mathbb{I} \end{pmatrix} \text{ is block-diagonal,} \tag{C.53}$$

and here $I(\rho_0, M)'$ is the FIM when the dual basis is taken as $(Q'\ T)$ where $Q' = Q + TC_1$ for $C_1 = -(I(\rho_0, M)_{BB})^{-1} I(\rho_0, M)_{BA}$ is a function of ρ_0 and M (Here $^{-1}$ means the pseudoinverse). Then let $Q'(\rho_0) := Q + TC_1^\star$ for $C_1^\star = -(I(\rho_0, M^\star)_{BB})^{-1} I(\rho_0, M^\star)_{BA}$.

Proof We first assume the FIM $I(\rho_0, M)$ is block-diagonal for any ρ_0, i.e., $I(\rho_0, M)_{AB} = 0$. Then we want to find a threshold on ε such that

$$\begin{aligned} \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \chi^2_M(\rho_{\theta,\varphi}\|\rho_0) &= \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi) \\ &= \min_{\|\theta\|_p=3\varepsilon} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi). \end{aligned} \tag{C.54}$$

Since the FIM $I(\rho_0, M)$ is block-diagonal,

$$\begin{aligned} \min_{\|\theta\|_p=3\varepsilon} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi) &= \min_{\|\theta\|_p=3\varepsilon} (\theta^\top I(\rho_0, M)_{AA}\theta + \varphi^\top I(\rho_0, M)_{BB}\varphi) \\ &= \min_{\|\theta\|_p=3\varepsilon} \theta^\top I(\rho_0, M)_{AA}\theta. \end{aligned} \tag{C.55}$$

Furthermore, if for all θ satisfying $\|\theta\|_p = 3\varepsilon$, $(\theta, 0) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$, then

$$\min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi) = \min_{\|\theta\|_p=3\varepsilon} \theta^\top I(\rho_0, M)_{AA}\theta, \tag{C.56}$$

because

$$\begin{aligned} \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi) &\le \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,0)^\top I(\rho_0, M)(\theta,0) \\ &= \min_{\|\theta\|_p=3\varepsilon} \theta^\top I(\rho_0, M)_{AA}\theta, \end{aligned} \tag{C.57}$$

and

$$\begin{aligned} \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi) &\ge \min_{\|\theta\|_p=3\varepsilon} (\theta,\varphi)^\top I(\rho_0, M)(\theta,\varphi) \\ &= \min_{\|\theta\|_p=3\varepsilon} \theta^\top I(\rho_0, M)_{AA}\theta. \end{aligned} \tag{C.58}$$

Therefore, under the assumption that $I(\rho_0, M)$ is block-diagonal, Eq. (C.54) holds if $\|\theta\|_p = 3\varepsilon$ implies $(\theta, 0) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$.

Next, we show any $I(\rho_0, M)$ can be made block-diagonal by properly choosing basis $\{Q_a, T_b\}$ for each ρ_0 and M. The basis transformation

$$\begin{pmatrix} Q' & T' \end{pmatrix} = \begin{pmatrix} Q + TC_1 & TC_2 \end{pmatrix} = \begin{pmatrix} Q & T \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ C_1 & C_2 \end{pmatrix} \tag{C.59}$$

corresponds to the new FIM

$$I(\rho_0, M)' = \begin{pmatrix} \mathbb{I} & C_1^\top \\ 0 & C_2^\top \end{pmatrix} \begin{pmatrix} I(\rho_0, M)_{AA} & I(\rho_0, M)_{AB} \\ I(\rho_0, M)_{BA} & I(\rho_0, M)_{BB} \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ C_1 & C_2 \end{pmatrix}. \tag{C.60}$$

Here we take $C_2 = \mathbb{I}$, which means

$$T' = T \tag{C.61}$$

and

$$C_1 = -(I(\rho_0, M)_{BB})^{-1} I(\rho_0, M)_{BA} \tag{C.62}$$

where $(I(\rho_0, M)_{BB})^{-1}$ is the pseudoinverse of $I(\rho_0, M)_{BB}$ on its support, is a solution of C_1 such that

$$I(\rho_0, M)'_{AB} = I(\rho_0, M)_{AB} C_2 + C_1^\top I(\rho_0, M)_{BB} C_2 = 0. \tag{C.63}$$

The corresponding choice of basis $Q' = Q + TC_1$ makes the corresponding FIM $I(\rho_0, M)'$ block-diagonal. It implies

$$\frac{1}{6\left\|\left(\|Q'_1\|_\infty, \ldots, \left\|Q'_{|A|}\right\|_\infty\right)\right\|_q} \tag{C.64}$$

is a threshold on ε below which Eq. (C.54) holds after the basis transformation. Below we fix Q' to be the above choice of basis (for different ρ_0 and M) such that $I(\rho_0, M)'$ is block-diagonal.

Finally, when

$$\varepsilon \le \eta := \inf_{\rho_0 \in \mathcal{S}_{1/2}} \frac{1}{6\left\|\left(\|Q'_1(\rho_0)\|_\infty, \ldots, \left\|Q'_{|A|}(\rho_0)\right\|_\infty\right)\right\|_q}, \tag{C.65}$$

where $Q'(\rho_0) = Q'$ that is defined above to make $I(\rho_0, M^\star)$ block-diagonal, which is a function of ρ_0, we have

$$\begin{aligned} &\inf_{M \in \mathcal{M}} \sup_{\substack{\rho_0 \in \mathcal{S}^\circ \\ (Q,T) \in \mathscr{B}(O)}} \max_{(\theta,\varphi) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \left((\theta, \varphi)^\top I(\rho_0, M)(\theta, \varphi)\right)^{-1} \\ &\le \inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\theta\|_p = 3\varepsilon} \left((\theta, \varphi)^\top I(\rho_0, M)(\theta, \varphi)\right)^{-1} \end{aligned} \tag{C.66}$$

where we use Lemma 15, and using Eq. (C.52), Lemma 13, Lemma 15 and Eq. (C.34), we have

$$\begin{aligned}
&\inf_{M\in\mathcal{M}} \sup_{\substack{\rho_0\in\mathcal{S}^\circ\\(Q,T)\in\mathscr{B}(O)}} \max_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \Big((\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)\Big)^{-1}\\
\geq{}& \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}_{1/2}} \max_{(\theta,\varphi)\in\mathscr{D}^{Q',T}_{3\varepsilon,p}(\rho_0)} \Big((\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi)\Big)^{-1}\\
\geq{}& \frac{1}{2} \sup_{\rho_0\in\mathcal{S}_{1/2}} \max_{(\theta,\varphi)\in\mathscr{D}^{Q',T}_{3\varepsilon,p}(\rho_0)} \Big((\theta,\varphi)^\top I(\rho_0,M^\star)'(\theta,\varphi)\Big)^{-1}\\
={}& \frac{1}{2} \sup_{\rho_0\in\mathcal{S}_{1/2}} \max_{\|\theta\|_p=3\varepsilon} \Big((\theta,\varphi)^\top I(\rho_0,M^\star)'(\theta,\varphi)\Big)^{-1}\\
={}& \frac{1}{2} \sup_{\rho_0\in\mathcal{S}_{1/2}} \min_{\|\alpha\|_q\leq 3\varepsilon} \Big(\alpha^\top I(\rho_0,M^\star)'^{-1}_{AA}\alpha\Big)^{-1} \qquad\text{(C.67)}\\
={}& \frac{1}{2} \sup_{\rho_0\in\mathcal{S}_{1/2}} \min_{\|\alpha\|_q\leq 3\varepsilon} \Big(\alpha^\top I(\rho_0,M^\star)^{-1}_{AA}\alpha\Big)^{-1}\\
={}& \frac{1}{2} \sup_{\rho_0\in\mathcal{S}_{1/2}} \max_{\|\theta\|_p=3\varepsilon} \Big((\theta,\varphi)^\top I(\rho_0,M^\star)(\theta,\varphi)\Big)^{-1}\\
\geq{}& \frac{1}{2} \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}_{1/2}} \max_{\|\theta\|_p=3\varepsilon} \Big((\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)\Big)^{-1}\\
\geq{}& \frac{1}{4} \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\|\theta\|_p=3\varepsilon} \Big((\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)\Big)^{-1}.
\end{aligned}$$

The theorem is then proved. Note that the above chain of inequalities holds when T is fixed and only the choice of Q is optimized over. ■

C.4. Few-copy measurements

We then consider protocols using ($c \geq 1$)-copy measurements in $\mathcal{M}_c$. We now show that for the many-versus-one distinguishing task, if we only care about ε^2 term, which is the case when ε is below a certain threshold, then the lower bound for these protocols can only achieve at most an $O(c)$ reduction from the single-copy protocol.

We first prove the following lemma, which will be used later in the proof of the theorem.

Lemma 17 (Minimax theorem) *Fix T in the dual basis. Let $\widetilde{\mathcal{D}}$ denote the set of all probability distributions of $(\rho_0,\tilde{Q},\theta,\varphi)$ over the set*

$$\{(\rho_0,\tilde{Q},\theta,\varphi)|\|\theta\|_p = 3\varepsilon, \theta\in\mathbb{R}^{|A|}, \varphi\in\mathbb{R}^{|B|}, \rho_0\in\mathcal{S}_{1/2}, \tilde{Q} = Q + TC_1, C_1\in\mathbb{R}^{|B|\times|A|}\}, \quad\text{(C.68)}$$

$\widetilde{\mathcal{D}}^T \subseteq \widetilde{\mathcal{D}}$ denote the set of all probability distributions of $(\rho_0,\tilde{Q},\theta,\varphi)$ over the set

$$\{(\rho_0,\tilde{Q},\theta,\varphi)|\rho_0\in\mathcal{S}_{1/2}, (\theta,\varphi)\in\mathscr{D}^{\tilde{Q},T}_{3\varepsilon,p}(\rho_0), \tilde{Q} = Q + TC_1, C_1\in\mathbb{R}^{|B|\times|A|}\}, \qquad\text{(C.69)}$$

$\widetilde{\mathcal{D}}'^T \subseteq \widetilde{\mathcal{D}}^T$ denote the set of all probability distributions of $(\rho_0,\tilde{Q},\theta,\varphi)$ over the set

$$\{(\rho_0,\tilde{Q},\theta,\varphi)|\rho_0\in\mathcal{S}_{1/2}, (\theta,\varphi)\in\mathscr{D}^{\tilde{Q},T}_{3\varepsilon,p}(\rho_0), \tilde{Q} = Q'(\rho_0)\}, \qquad\text{(C.70)}$$

where $Q'(\rho_0)$ was defined in the statement of Theorem 16. We have

$$\sup_{M\in\mathcal{M}}\inf_{\rho_0\in\mathcal{S}_{1/2},\tilde{Q}}\min_{\|\theta\|_p=3\varepsilon}(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi)=\inf_{\pi\in\widetilde{\mathcal{D}}}\sup_{M\in\mathcal{M}}\mathbb{E}_\pi(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi), \tag{C.71}$$

$$\sup_{M\in\mathcal{M}}\inf_{\rho_0\in\mathcal{S}_{1/2},\tilde{Q}}\min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)}(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi)=\inf_{\pi\in\widetilde{\mathcal{D}}^T}\sup_{M\in\mathcal{M}}\mathbb{E}_\pi(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi), \tag{C.72}$$

$$\sup_{M\in\mathcal{M}}\inf_{\rho_0\in\mathcal{S}_{1/2}}\min_{(\theta,\varphi)\in\mathscr{D}^{Q',T}_{3\varepsilon,p}(\rho_0)}(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi)=\inf_{\pi\in\widetilde{\mathcal{D}}'^T}\sup_{M\in\mathcal{M}}\mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi). \tag{C.73}$$

Here we use $\mathbb{E}_\pi$ as a short-hand of $\mathbb{E}_{(\rho_0,\tilde{Q},\theta,\varphi)\sim\pi}$ and $\tilde{I}$ as a short-hand of the FIM when the dual basis is taken as $(\tilde{Q}\;T)$. Here $\tilde{Q}$ implicitly belongs to the set $\{\tilde{Q}|\tilde{Q}=Q+TC_1,\forall C_1\in\mathbb{R}^{|B|\times|A|}\}$ which is a function of $\{O_i\}_{i=1}^m$ only.

Proof We first apply the following result (Chiribella et al., 2007) which states that any POVM $M(S)$ for $S\subseteq\Omega$ can be decomposed as

$$M(S)=\int_{\mathcal{X}}\mathrm{d}xp(x)E^{(x)}(S), \tag{C.74}$$

where $x\in\mathcal{X}$ is a suitable random variable, $p(x)$ a probability density and $E^{(x)}$ is a POVM with finite support, i.e.

$$E^{(x)}(S)=\sum_{i=1}^{d^2}\mathbb{I}[\omega_i\in S]E_i, \tag{C.75}$$

where $\{\omega_i\}_{i=1}^{d^2}\subseteq\Omega$, $\mathbb{I}[\cdot]$ is the indicator function and E_i is a POVM with (at most) d^2 outcomes. Since $I(\rho_0,M)=\int_{\mathcal{X}}\mathrm{d}xp(x)I(\rho_0,E^{(x)})$ from the definition of FIM, we have

$$\begin{aligned}\sup_{M\in\mathcal{M}}(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)&\le\sup_{p(x)}\int_{\mathcal{X}}\mathrm{d}xp(x)\sup_{E^{(x)}\in\mathcal{M}^{[d^2]}}(\theta,\varphi)^\top I(\rho_0,E^{(x)})(\theta,\varphi)\\&=\sup_{E\in\mathcal{M}^{[d^2]}}(\theta,\varphi)^\top I(\rho_0,E)(\theta,\varphi),\end{aligned} \tag{C.76}$$

and then $\sup_{M\in\mathcal{M}}(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)=\sup_{M\in\mathcal{M}^{[d^2]}}(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)$. Similarly,

$$\sup_{M\in\mathcal{M}}\mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)=\sup_{M\in\mathcal{M}^{[d^2]}}\mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi). \tag{C.77}$$

Next we show,

$$\inf_{\pi\in\widetilde{\mathcal{D}}}\sup_{M\in\mathcal{M}}\mathbb{E}_\pi(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi)=\sup_{M\in\mathcal{M}}\inf_{\pi\in\widetilde{\mathcal{D}}}\mathbb{E}_\pi(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi), \tag{C.78}$$

which proves Eq. (C.71). Since the other direction is trivial and thanks to Eq. (C.77), we only need to show

$$\inf_{\pi\in\widetilde{\mathcal{D}}}\sup_{M\in\mathcal{M}^{[d^2]}}\mathbb{E}_\pi(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi)\le\sup_{M\in\mathcal{M}^{[d^2]}}\inf_{\pi\in\widetilde{\mathcal{D}}}\mathbb{E}_\pi(\theta,\varphi)^\top\tilde{I}(\rho_0,M)(\theta,\varphi). \tag{C.79}$$

This can be proven using Sion's minimax theorem (Sion, 1958) to exchange the order of $\inf$ and $\sup$. It states that $\inf_{x\in X}\sup_{y\in Y} f(x,y) = \sup_{y\in Y}\inf_{x\in X} f(x,y)$ if X is a convex and compact subset of a linear topological space, Y is a convex subset of a linear topological space, $x \mapsto f(x,y)$ is continuous and convex, $y \mapsto f(x,y)$ is continuous and concave. First, to exchange the order of $\inf_\pi$ and $\sup_M$, we notice $\mathcal{M}^{[d^2]}$ is a convex, compact subset of a linear topological space, $\widetilde{\mathcal{D}}$ is a convex subset of a linear topological space, $\mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)$ is convex in M and concave in π. Eq. (C.72) and Eq. (C.73) can be proven similarly. Note that here when we say $\mathcal{M}^{[d^2]}$ is convex and $\mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi)$ is convex in M, we implicitly use the following definition of convex combination:

$$\lambda\{M_s\}_{s=1}^{d^2} + (1-\lambda)\{M_s'\}_{s=1}^{d^2} = \{\lambda M_s + (1-\lambda)M_s'\}_{s=1}^{d^2}. \tag{C.80}$$

■

Below we derive an important property implied by Lemma 17 that will be used in the proof later. Assume $\pi^\star$ be a nearly optimal distribution for some fixed T such that

$$\begin{aligned} &\sup_{M\in\mathcal{M}} \mathbb{E}_{\pi^\star}(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi) \\ &\le 2\sup_{M\in\mathcal{M}}\inf_{\pi\in\widetilde{\mathcal{D}}'^T} \mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi) \\ &\le 4\inf_{\pi\in\widetilde{\mathcal{D}}'^T} \mathbb{E}_\pi(\theta,\varphi)^\top I(\rho_0,M^\star)'(\theta,\varphi) \\ &= 4\inf_{\pi\in\widetilde{\mathcal{D}}'^T} \mathbb{E}_\pi\Big[\theta^\top I(\rho_0,M^\star)'_{AA}\theta + \varphi^\top I(\rho_0,M^\star)'_{BB}\varphi\Big]. \end{aligned} \tag{C.81}$$

Define for any $c_2 > 0$,

$$\Lambda_{c_2}\pi(\rho_0,\tilde{Q},\theta,\varphi) := c_2\pi(\rho_0,\tilde{Q},\theta,c_2\varphi). \tag{C.82}$$

Then since I' is block-diagonal we have for any $c_2 \ge 1$,

$$\begin{aligned} &\sup_{M\in\mathcal{M}} \mathbb{E}_{\Lambda_{c_2}\pi^\star}(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi) \\ &\le \sup_{M\in\mathcal{M}} \mathbb{E}_{\pi^\star}(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi) \\ &\le 4\inf_{\pi\in\widetilde{\mathcal{D}}'^T} \mathbb{E}_\pi\Big[\theta^\top I(\rho_0,M^\star)'_{AA}\theta + \varphi^\top I(\rho_0,M^\star)'_{BB}\varphi\Big]. \end{aligned} \tag{C.83}$$

To conclude, $\Lambda_{c_2}\pi^\star$ for any $c_2 \ge 1$ is also a nearly optimal distribution satisfying

$$\begin{aligned} \inf_{\pi\in\widetilde{\mathcal{D}}}\sup_{M\in\mathcal{M}} \mathbb{E}_\pi(\theta,\varphi)^\top \tilde{I}(\rho_0,M)(\theta,\varphi) &\le \sup_{M\in\mathcal{M}} \mathbb{E}_{\Lambda_{c_2}\pi^\star}(\theta,\varphi)^\top I(\rho_0,M)'(\theta,\varphi) \\ &\le 16\inf_{\pi\in\widetilde{\mathcal{D}}}\sup_{M\in\mathcal{M}} \mathbb{E}_\pi(\theta,\varphi)^\top \tilde{I}(\rho_0,M)(\theta,\varphi), \end{aligned} \tag{C.84}$$

which can be seen from the proof of Theorem 16.

Finally, we have the following theorem.

Theorem 18 **(Lower bound for Problem 3(3') and p-norm error)** *Using (possibly adaptive) measurement strategy with c-copy measurements, consider the many-versus-one distinguishing tasks above with any $\rho_{\theta,\varphi}$ well-defined and $(\theta, \varphi) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)$. The sample complexity required to solve this task is*

$$N = \Omega\left(\frac{\Gamma^{\mathrm{ob}}_p(\{O_i\}_{i=1}^m)}{c\varepsilon^2}\right), \tag{C.85}$$

for any $\varepsilon \leq \min\{\eta^{\mathrm{ob}}, \eta^{\mathrm{ob}}_c\}$ where

$$\eta^{\mathrm{ob}}_c := \min\left\{\frac{1}{3ca_{\max}\sqrt{\Gamma^{\mathrm{ob}}_p(\{O_i\}_{i=1}^m)}}, \frac{1}{12c\sqrt{a_{\max}}}\right\}, \tag{C.86}$$

with

$$\begin{aligned} a_{\max} &:= \max_{\rho_0\in\mathcal{S}_{1/2},\|\theta\|_p=1} a_{\rho_0}(\theta), \\ a_{\rho_0}(\theta) &:= \frac{\mathrm{tr}\big((\theta\cdot\boldsymbol{Q}')\rho_0^{-1}(\theta\cdot\boldsymbol{Q}')\big)}{d^2} = \theta^\top G^{(\rho_0)}\theta, \quad G^{(\rho_0)}_{ij} := \frac{\mathrm{tr}(Q'_i\rho_0^{-1}Q'_j)}{d^2}, \end{aligned} \tag{C.87}$$

Q' is a function of ρ_0 as defined in Theorem 16, and $p \in [1, \infty]$.

Proof

Note that the sample complexity bound from Eq. (B.17) is now $\Omega(c/\delta_{\mathcal{M}_c}(O))$ where

$$\delta_{\mathcal{M}_c}(O) = \sup_{M\in\mathcal{M}_c} \inf_{\substack{\rho_0\in\mathcal{S}^\circ \\ (Q,T)\in\mathscr{B}(O)}} \min_{(\theta,\varphi)\in\mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \chi^2_M\left(\rho_{\theta,\varphi}^{\otimes c}\|\rho_0^{\otimes c}\right). \tag{C.88}$$

We denote the c-copy POVM as $M = \{M_s\}_s$, and the classical distributions

$$q_s = \mathrm{tr}(M_s\rho_0^{\otimes c}), \qquad p_s(\theta,\varphi) = \mathrm{tr}\left(M_s\rho_{\theta,\varphi}^{\otimes c}\right), \tag{C.89}$$

the χ^2 divergence induced by M is

$$\chi^2_M(\rho_{\theta,\varphi}^{\otimes c}\|\rho_0^{\otimes c}) = \sum_s \frac{\big(p_s(\theta,\varphi) - q_s\big)^2}{q_s}. \tag{C.90}$$

We expand $p_s(\theta,\varphi)$ around ρ_0. Using the notation

$$\rho_{\theta,\varphi} = \rho_0 + \frac{(\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})}{d}, \tag{C.91}$$

where $\boldsymbol{Q} = (Q_1, ..., Q_{|A|})$ and $\boldsymbol{T} = (T_1, ..., T_{|B|})$, the multinomial expansion over subsets $S \subseteq [c]$ is

$$\rho_{\theta,\varphi}^{\otimes c} = \sum_{S\subseteq[c]} \left(\frac{(\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})}{d}\right)^{(S)} \otimes \rho_0^{([c]\setminus S)}, \tag{C.92}$$

where $(\cdot)^{(S)}$ denotes placing the operator on the tensor factors indexed by S and identity elsewhere. Hence

$$p_s(\theta,\varphi) - q_s = \sum_{\varnothing \neq S \subseteq [c]} \operatorname{tr}\left(M_s \left((\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T}) \right)^{(S)} \otimes \rho_0^{([c]\setminus S)} \right) \frac{1}{d^{|S|}}, \tag{C.93}$$

Therefore,

$$\chi^2_M(\rho_{\theta,\varphi}^{\otimes c} \| \rho_0^{\otimes c}) = \sum_s \frac{1}{\operatorname{tr}(M_s \rho_0^{\otimes c})} \left(\sum_{\varnothing \neq S \subseteq [c]} \frac{\operatorname{tr}\left(M_s \left((\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T}) \right)^{(S)} \otimes \rho_0^{([c]\setminus S)} \right)}{d^{|S|}} \right)^2 . \tag{C.94}$$

For some fixed T in the dual basis, we pick some $\pi^\star \in \widetilde{\mathcal{D}}^T$ satisfying Eq. (C.84). Let $(\bar{\theta},\bar{\varphi}) := (\theta,\varphi)/d$. Since

$$\begin{aligned}
&\chi^2_M(\rho_{\theta,\varphi}^{\otimes c} \| \rho_0^{\otimes c}) \\
&= \sum_s \frac{1}{\operatorname{tr}(M_s \rho_0^{\otimes c})} \left(\sum_{\varnothing \neq S \subseteq [c]} \frac{\operatorname{tr}\left(M_s \left((\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T}) \right)^{(S)} \rho_0^{([c]\setminus S)} \right)}{d^{|S|}} \right)^2 \\
&\leq \left(\sum_{\varnothing \neq S \subseteq [c]} \frac{1}{w_S} \sum_s \frac{\operatorname{tr}\left(M_s \left((\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T}) \right)^{(S)} \rho_0^{([c]\setminus S)} \right)^2}{\operatorname{tr}(M_s \rho_0^{\otimes c}) d^{2|S|}} \right) \left(\sum_{\varnothing \neq S \subseteq [c]} w_S \right) \\
&\leq \left(\sum_{\varnothing \neq S \subseteq [c]} \sqrt{ \sum_s \frac{\operatorname{tr}\left(M_s \left((\bar{\theta},\bar{\varphi})\cdot(\boldsymbol{Q},\boldsymbol{T}) \right)^{(S)} \rho_0^{([c]\setminus S)} \right)^2}{\operatorname{tr}(M_s \rho_0^{\otimes c})} } \right)^2 ,
\end{aligned} \tag{C.95}$$

where the third step used Cauchy-Schwartz and the last step optimizes over all choices of w_S, we have for any fixed T,

$$\begin{aligned}
\sqrt{\delta_{\mathcal{M}_c}(O)} &\leq \sup_{M \in \mathcal{M}_c} \inf_{\rho_0 \in \mathcal{S}_{1/2}} \min_{(\theta,\varphi) \in \mathscr{D}^{Q,T}_{3\varepsilon,p}(\rho_0)} \sqrt{\chi^2_M\left(\rho_{\theta,\varphi}^{\otimes c} \| \rho_0^{\otimes c} \right)} \\
&\leq \sup_{M \in \mathcal{M}_c} \inf_{\pi \in \widetilde{\mathcal{D}}^T} \mathbb{E}_\pi \sqrt{\chi^2_M\left(\tilde{\rho}_{\theta,\varphi}^{\otimes c} \| \rho_0^{\otimes c} \right)} \\
&\leq \sup_{M \in \mathcal{M}_c} \inf_{\pi \in \widetilde{\mathcal{D}}^T} \mathbb{E}_\pi \sum_{\varnothing \neq S \subseteq [c]} \sqrt{ \sum_s \frac{\operatorname{tr}\left(M_s \left((\bar{\theta},\bar{\varphi})\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T}) \right)^{(S)} \rho_0^{([c]\setminus S)} \right)^2}{\operatorname{tr}(M_s \rho_0^{\otimes c})} } \\
&\leq \sup_{M \in \mathcal{M}_c} \mathbb{E}_{\pi^\star} \sum_{\varnothing \neq S \subseteq [c]} \sqrt{ \sum_s \frac{\operatorname{tr}\left(M_s \left((\bar{\theta},\bar{\varphi})\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T}) \right)^{(S)} \rho_0^{([c]\setminus S)} \right)^2}{\operatorname{tr}(M_s \rho_0^{\otimes c})} } \\
&\leq \Delta_1^{T,\pi^\star} + \Delta_2^{T,\pi^\star},
\end{aligned} \tag{C.96}$$

where $\tilde{\rho}_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a\tilde{Q}_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b$,

$$\Delta_1^{T,\pi^\star} := \sum_{|S|=1}\sqrt{\mathbb{E}_{\pi^\star}\sup_{M\in\mathcal{M}_c}\sum_s\frac{\operatorname{tr}\left(M_s\left((\bar{\theta},\bar{\varphi})\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T})\right)^{(S)}\rho_0^{([c]\setminus S)}\right)^2}{\operatorname{tr}(M_s\rho_0^{\otimes c})}}, \tag{C.97}$$

and

$$\Delta_2^{T,\pi^\star} := \sum_{|S|>1}\sqrt{\mathbb{E}_{\pi^\star}\sup_{M\in\mathcal{M}_c}\sum_s\frac{\operatorname{tr}\left(M_s\left((\bar{\theta},\bar{\varphi})\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T})\right)^{(S)}\rho_0^{([c]\setminus S)}\right)^2}{\operatorname{tr}(M_s\rho_0^{\otimes c})}}. \tag{C.98}$$

Note that the above inequality still holds when replacing $\pi^\star$ with any other distribution.

We first prove the lowest-order term is bounded as expected. Define, from any c-copy POVM element M_s, a single-copy operator

$$G_s := \operatorname{tr}_{[c]\setminus\{1\}}\big((\mathbb{I}\otimes\rho_0^{(c-1)})M_s\big). \tag{C.99}$$

We can verify that $\{G_s\}_s$ is a valid single-copy measurement.

- Positivity: $G_s \geq 0$ since $G_s^\dagger = \operatorname{tr}_{[c]\setminus\{1\}}\big(M_s(\mathbb{I}\otimes\rho_0^{(c-1)})\big) = \operatorname{tr}_{[c]\setminus\{1\}}\big((\mathbb{I}\otimes\rho_0^{(c-1)})M_s\big) = G_s$ by cyclicity and $\operatorname{tr}(AG_s) = \operatorname{tr}(M_s(A\otimes\rho_0^{\otimes(c-1)}))$ for any operator A, implying G_s is positive semidefinite.
- Completeness: $\sum_s G_s = \operatorname{tr}_{[c]\setminus\{1\}}(\mathbb{I}\otimes\rho_0^{\otimes(c-1)}) = \mathbb{I}$ (because $\operatorname{tr}(\rho_0)=1$). So $G=\{G_s\}_s$ is a valid single-copy POVM.
- Crucial identities: For every x and every observable A acting on the first copy,

$$\frac{\operatorname{tr}\left(M_s(A^{(1)}\otimes\rho_0^{\otimes(c-1)})\right)^2}{\operatorname{tr}(M_s\rho_0^{\otimes c})} = \frac{\operatorname{tr}(AG_s)^2}{\operatorname{tr}(G_s\rho_0)}. \tag{C.100}$$

Now, we note that $A = (\bar{\theta},\bar{\varphi})\cdot(\boldsymbol{Q},\boldsymbol{T})$, the $|S|=1$ functional equals the single-copy functional for G with base state ρ_0:

$$\sum_s\frac{\operatorname{tr}\left(M_s\left((\bar{\theta},\bar{\varphi})\cdot(\boldsymbol{Q},\boldsymbol{T})\right)^{(1)}\rho_0^{(c-1)}\right)^2}{\operatorname{tr}(M_s\rho_0^{\otimes c})} = \sum_s\frac{\operatorname{tr}\left(G_s\left((\bar{\theta},\bar{\varphi})\cdot(\boldsymbol{Q},\boldsymbol{T})\right)\right)^2}{\operatorname{tr}(G_s\rho_0)} = (\theta,\varphi)^\top I(\rho_0,G)(\theta,\varphi). \tag{C.101}$$

Here $^{(1)}$ means operator acting on the first qudit. The mapping sends any $M\in\mathcal{M}_c$ to some $G\in\mathcal{M}$ that is a single-copy POVM. Furthermore, any single-copy POVM G can be written in the form of Eq. (C.99) when taking $M_s = G_s\otimes\mathbb{I}^{(c-1)}$. Therefore,

$$\sup_{M\in\mathcal{M}_c}\sum_s\frac{\operatorname{tr}\left(M_s\left((\bar{\theta},\bar{\varphi})\cdot(\boldsymbol{Q},\boldsymbol{T})\right)^{(1)}\rho_0^{\otimes(c-1)}\right)^2}{\operatorname{tr}(M_s\rho_0^{\otimes c})} = \sup_{M\in\mathcal{M}}(\theta,\varphi)^\top I(\rho_0,M)(\theta,\varphi). \tag{C.102}$$

Due to Eq. (C.84), we have

$$\frac{(3\varepsilon)^2}{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)} \leq \frac{(\Delta_1^{T,\Lambda_{c_2}\pi^\star})^2}{c} \leq \frac{16(3\varepsilon)^2}{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}, \tag{C.103}$$

for any $c_2 \geq 1$.

It remains to derive the threshold on ε below which we can ignore the contributions from higher-order terms. Since

$$\sqrt{\delta_{\mathcal{M}_c}(O)} \leq \Delta_1^{T,\Lambda_{c_2}\pi^\star} + \Delta_2^{T,\Lambda_{c_2}\pi^\star} \tag{C.104}$$

holds for any $c_2 \geq 1$, it is sufficient to show

$$\inf_{c_2\geq 1} \Delta_2^{T,\Lambda_{c_2}\pi^\star} \leq c\sqrt{\frac{(3\varepsilon)^2}{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}}. \tag{C.105}$$

Fix a subset S with $|S| = k$. Set

$$N_s := ((\rho_0)^{1/2})^{\otimes c} M_s ((\rho_0)^{1/2})^{\otimes c} \qquad \Big(\Rightarrow \sum_s N_s = \rho_0^{\otimes c}, \mathrm{tr}(N_s) = \mathrm{tr}(M_s\rho_0^{\otimes c})\Big), \tag{C.106}$$

and define

$$B_S := \bigotimes_{i\in S} \left(\rho_0^{-1/2} \left((\bar\theta,\bar\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})\right)^{(i)} \rho_0^{-1/2}\right) \otimes \bigotimes_{i\notin S} \mathbb{I}. \tag{C.107}$$

Then

$$\mathrm{tr}\Big(M_s \left((\bar\theta,\bar\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})\right)^{(S)} \otimes \rho_0^{([c]\setminus S)}\Big) = \mathrm{tr}\big(N_s B_S\big), \qquad \mathrm{tr}(M_s\rho_0^{\otimes c}) = \mathrm{tr}(N_s). \tag{C.108}$$

By Cauchy–Schwarz in the Hilbert–Schmidt inner product,

$$\sum_s \frac{\mathrm{tr}(N_s B_S)^2}{\mathrm{tr}(N_s)} \leq \mathrm{tr}\left(B_S\left(\sum_s N_s\right)B_S\right) = \mathrm{tr}\big(B_S\rho_0^{\otimes c}B_S\big). \tag{C.109}$$

Evaluating the right-hand side factorizes over tensor slots, giving

$$\mathrm{tr}\big(B_S\rho_0^{\otimes c}B_S\big) = \Big[\mathrm{tr}\left(\left((\bar\theta,\bar\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})\right)\rho_0^{-1}\left((\bar\theta,\bar\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})\right)\right)\Big]^k. \tag{C.110}$$

Therefore, we have

$$\begin{aligned}
&\inf_{c_2\geq 1}\Delta_2^{T,\Lambda_{c_2}\pi^\star} \\
&= \inf_{c_2\geq 1}\sum_{|S|>1}\sqrt{\mathbb{E}_{\pi^\star}\sup_{M\in\mathcal{M}_c}\sum_s \frac{\mathrm{tr}\left(M_s\left((\bar\theta,\bar\varphi)\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T})\right)^{(S)}\rho_0^{([c]\setminus S)}\right)^2}{\mathrm{tr}(M_s\rho_0^{\otimes c})}} \\
&\leq \inf_{c_2\geq 1}\sum_{k=2}^{c}\binom{c}{k}\frac{1}{d^k}\mathbb{E}_{\Lambda_{c_2}\pi^\star}\Big[\mathrm{tr}\left(\left((\bar\theta,\bar\varphi)\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T})\right)\rho_0^{-1}\left((\bar\theta,\bar\varphi)\cdot(\tilde{\boldsymbol{Q}},\boldsymbol{T})\right)\right)\Big]^{k/2} \\
&\leq \sum_{k=2}^{c}\binom{c}{k}\frac{1}{d^k}\mathbb{E}_{\mathscr{P}^\star}\Big[\mathrm{tr}\left(\left((\bar\theta)\cdot(\boldsymbol{Q}')\right)\rho_0^{-1}\left((\bar\theta)\cdot(\boldsymbol{Q}')\right)\right)\Big]^{k/2},
\end{aligned} \tag{C.111}$$

where $\mathbb{E}_{\mathscr{P}^\star} = \mathbb{E}_{(\rho_0,\tilde{Q},\varphi)\sim\mathscr{P}^\star}$. Here we use, as $c_2 \to \infty$,

$$\lim_{c_2\to\infty} \Lambda_{c_2}\pi^\star(\rho_0,\tilde{Q},\theta,\varphi) = \mathscr{P}^\star(\rho_0,\theta)\delta(\varphi)\delta(\tilde{Q}-Q'(\rho_0)), \tag{C.112}$$

where $\mathscr{P}^\star(\rho_0,\theta) = \int \pi^\star(\rho_0,\tilde{Q},\theta,\varphi)\mathrm{d}\varphi\mathrm{d}\tilde{Q}$ and $\delta(\cdot)$ is the delta function. The $\delta(\tilde{Q}-Q'(\rho_0))$ part stems from the definition of $\widetilde{\mathcal{D}}'^T$ and the $\delta(\varphi)$ part is because infinite squeezing in the domain of φ collapses the distribution into a delta function. Define the quadratic form

$$\begin{aligned} a_{\rho_0}(\theta) &:= \frac{\operatorname{tr}\Big((\theta\cdot\boldsymbol{Q}')\,\rho_0^{-1}\,(\theta\cdot\boldsymbol{Q}') \Big)}{d^2} = \theta^\top G^{(\rho_0)}\theta, \\ G^{(\rho_0)}_{ij} &:= \frac{\operatorname{tr}(Q'_i\rho_0^{-1}Q'_j)}{d^2}. \end{aligned} \tag{C.113}$$

With $x = \sqrt{a_{\rho_0}(\theta)}$,

$$\inf_{c_2\ge 1} \Delta_2^{T,\Lambda_{c_2}\pi^\star} \le \mathbb{E}_{\mathscr{P}^\star}[(1+x)^c - 1 - cx]. \tag{C.114}$$

Using $(1+x)^c \le e^{cx}$ and $e^t \le 1 + t + \frac{t^2}{2}e^t$ for $t \ge 0$,

$$\frac{\inf_{c_2\ge 1}\Delta_2^{T,\Lambda_{c_2}\pi^\star}}{c\sqrt{\frac{(3\varepsilon)^2}{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}}} \le \frac{\mathbb{E}_{\mathscr{P}^\star}[(cx)^2e^{cx}]}{2c\sqrt{\frac{(3\varepsilon)^2}{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}}}. \tag{C.115}$$

We define

$$a_{\max} := \max_{\rho_0\in\mathcal{S}_{1/2},\|\theta\|_p=1} a_{\rho_0}(\theta), \tag{C.116}$$

where we explicitly indicate the dependence of $a_{\max}$ on Q'. Therefore, with $x \le (3\varepsilon)\sqrt{a_{\max}}$, we get

$$\left(\frac{\inf_{c_2\ge 1}\Delta_2^{T,\Lambda_{c_2}\pi^\star}}{c\sqrt{\frac{(3\varepsilon)^2}{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}}}\right)^2 \le \frac{c^2(3\varepsilon)^4a_{\max}^2\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)e^{6c\varepsilon\sqrt{a_{\max}}}}{4(3\varepsilon)^2}. \tag{C.117}$$

To ensure the above is upper bounded by 1, it suffices to impose e.g. $6c\varepsilon\sqrt{a_{\max}} \le 1/2$ and $\frac{c^2(3\varepsilon)^4a_{\max}^2\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{4(3\varepsilon)^2} \le 1/4$, which implies

$$\varepsilon \le \min\left\{\frac{1}{3ca_{\max}\sqrt{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}}, \frac{1}{12c\sqrt{a_{\max}}}\right\} \tag{C.118}$$

is sufficient. ■

Appendix D. Lower bounds for unbiased, bounded estimation

In this section, we use the CR bound to prove a lower bound on the sample complexity required to solve Problem 2, i.e., estimation of parameters with p-norm error. Since the CR bound applies to only unbiased estimators, the sample complexity lower bound is also restricted to unbiased estimators. However, the bounds applies to general learning and estimation with p-norm error that is not restricted to the oblivious cases.

D.1. Single-copy measurements

Theorem 19 (Lower bound for estimation with p-norm error w/ bounded and unbiased estimators) *Using the adaptive measurement strategy with single-copy measurements, the sample complexity of ρ required to obtain a bounded, unbiased estimator of θ in Problem 2 is*

$$N = \Omega\left(\inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \frac{\left\|\operatorname{diag}\left((I(\rho_0, M)^{-1})_{AA}\right)^{1/2}\right\|_p^2}{\varepsilon^2 \log\left(m^{1/p}/\varepsilon\right)}\right) = \Omega\left(\frac{\Gamma_p(\{O_i\}_{i=1}^m)}{\varepsilon^2 \log\left(m^{1/p}/\varepsilon\right)}\right), \tag{D.1}$$

for any $\varepsilon > 0$ and $p \in [2, \infty]$, where $I(\rho_0, M) = I(\rho_{\theta,\varphi}, M)|_{\theta=\varphi=0}$, $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b$, and $\operatorname{diag}(\cdot)$ represents a diagonal matrix whose diagonal entries are those of $(\cdot)$. Here "bounded" means the values of estimators $\hat{\theta}_i$ are always away from true values by a constant.

Proof The proof consists of three steps.

1. First, we will show that for a fixed non-adaptive single-copy measurement strategy $M^{\otimes N_1}$, we need a sample complexity of

$$N_1 = \Omega\left(\sup_{\rho_0\in\mathcal{S}^\circ} \frac{\left\|\operatorname{diag}\left((I(\rho_0, M)^{-1})_{AA}\right)^{1/2}\right\|_p^2}{\xi^2}\right) \tag{D.2}$$

to construct an unbiased estimator $\hat{\theta}$ that achieves a p-average root-MSE smaller than ξ, i.e.,

$$\left(\sum_{i=1}^m \mathbb{E}[(\hat{\theta}_i - \theta_i)^2]^{p/2}\right)^{1/p} < \xi. \tag{D.3}$$

Given N_1 copies of parameterized quantum state

$$\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b, \tag{D.4}$$

where ρ_0 is known, $\{\theta_a\}_{a\in A}$ are to be estimated and $\{\varphi_b\}_{b\in B}$ are unknown (i.e. nuisance parameters Suzuki et al. (2020)), the CR bound states for any unbiased estimator $(\hat{\theta}^{(N_1)}, \varphi^{(N_1)})$,

$$V(\rho_{\theta,\varphi}, M^{\otimes N_1}, (\hat{\theta}^{(N_1)}, \varphi^{(N_1)})) \succeq \frac{1}{N_1} I(\rho_{\theta,\varphi}, M)^{-1}, \tag{D.5}$$

where $V(\rho_{\theta,\varphi}, M^{\otimes N_1}, (\hat{\theta}^{(N_1)}, \varphi^{(N_1)}))$ and $I(\rho_{\theta,\varphi}, M)$ are the MSEM and the FIM with respect to both parameters $\{\theta_a\}_{a\in A}$ and $\{\varphi_b\}_{b\in B}$. We can take the upper left blocks of the matrices that only involve entries in A, which gives

$$V(\rho_{\theta,\varphi}, M^{\otimes N_1}, \hat{\theta}^{(N_1)}) \succeq \frac{1}{N_1}(I(\rho_{\theta,\varphi}, M)^{-1})_{AA}, \tag{D.6}$$

Furthermore, we have

$$\operatorname{tr}(\operatorname{diag}(V(\rho_{\theta,\varphi}, M^{\otimes N_1}, \hat{\theta}^{(N_1)}))^{p/2}) = \sum_{i=1}^{m} \mathbb{E}[(\hat{\theta}_i - \theta_i)^2]^{p/2}. \tag{D.7}$$

That implies when Eq. (D.3) holds,

$$\xi \geq \frac{1}{\sqrt{N_1}} \operatorname{tr}\left(\operatorname{diag}\left((I(\rho_{\theta,\varphi}, M)^{-1})_{AA}\right)^{p/2}\right)^{1/p}. \tag{D.8}$$

Since we would like the above to hold for arbitrary (θ, φ) such that $\rho_{\theta,\varphi}$ is well defined. Let

$$\mathscr{D}(\rho_0) := \{(\theta, \varphi) | \rho_{\theta,\varphi} \text{ is a full-rank density matrix}\}, \tag{D.9}$$

which is an open set in $\mathbb{R}^{d^2-1}$. Then

$$\begin{aligned} \xi &\geq \frac{1}{\sqrt{N_1}} \sup_{(\theta,\varphi)\in\mathscr{D}(\rho_0)} \operatorname{tr}\left(\operatorname{diag}\left((I(\rho_{\theta,\varphi}, M)^{-1})_{AA}\right)^{p/2}\right)^{1/p} \\ &= \frac{1}{\sqrt{N_1}} \sup_{\rho_0\in\mathcal{S}^\circ} \operatorname{tr}\left(\operatorname{diag}\left((I(\rho_0, M)^{-1})_{AA}\right)^{p/2}\right)^{1/p}, \end{aligned} \tag{D.10}$$

where the last equality holds because choosing a specific value of (θ, φ) is equivalent to replacing the original ρ_0 with $\rho_{\theta,\varphi}$ and then setting its value to be zero.

2. Next, we show for $p \in [2, \infty]$, if an unbiased, bounded estimator $\hat{\theta}$ satisfies

$$\left\|\hat{\theta} - \theta\right\|_p = \left(\sum_{i=1}^{m} |\hat{\theta}_i - \theta_i|^p\right)^{1/p} < \varepsilon, \tag{D.11}$$

with probability $> 1 - \delta$, and

$$|\hat{\theta}_i - \theta_i| \leq u, \tag{D.12}$$

for all i and some constant u, as required by Problem 2, then we can construct another unbiased estimator $\hat{\theta}_K$ that satisfies Eq. (D.3) with $\xi = 4\sqrt{2}\varepsilon$ and overhead $K = O\big(\log\big(m^{1/p}/\varepsilon\big)\big)$, (i.e. $\hat{\theta}_K$ uses K samples of the estimator $\hat{\theta}$). Without loss of generality, we assume $\delta < 1/3$.

Taking K independent samples of $\hat{\theta}$, we define

$$\hat{\theta}_K := \text{Median}_p(\hat{\theta}^{[1]}, \ldots, \hat{\theta}^{[K]}) = \arg\min_{\tilde{\theta}} \sum_{\ell=1}^{K} \left\|\hat{\theta}^{[\ell]} - \tilde{\theta}\right\|_p, \tag{D.13}$$

where $\hat{\theta}^{[\ell]}$ are the ℓ-th sample of $\hat{\theta}$ and $\text{Median}_p(\hat{\theta}^{[1]},\dots,\hat{\theta}^{[K]})$ is the geometric median. One property of the median estimator is if $\|\hat{\theta}^{[\ell]}-\mathbb{E}[\hat{\theta}]\|_p<\varepsilon$ holds for ratio $\kappa>1/2$ of all ℓ, we must have $\|\hat{\theta}_K-\mathbb{E}[\hat{\theta}]\|_p<\frac{2\kappa}{2\kappa-1}\varepsilon$. It can be seen by noting that

$$\begin{aligned}\|\hat{\theta}_K-\hat{\theta}^{[\ell]}\|_p-\|\hat{\theta}^{[\ell]}-\mathbb{E}[\hat{\theta}]\|_p&>\|\hat{\theta}_K-\mathbb{E}[\hat{\theta}]\|_p-2\varepsilon,\quad\text{for }\ell\text{ satisfying }\|\hat{\theta}^{[\ell]}-\mathbb{E}[\hat{\theta}]\|_p<\varepsilon;\\ \|\hat{\theta}_K-\hat{\theta}^{[\ell]}\|_p-\|\hat{\theta}^{[\ell]}-\mathbb{E}[\hat{\theta}]\|_p&\geq-\|\hat{\theta}_K-\mathbb{E}[\hat{\theta}]\|_p,\quad\text{otherwise.}\end{aligned}\tag{D.14}$$

Summing over all ℓ, we have $0>(2\kappa-1)\|\hat{\theta}_K-\mathbb{E}[\hat{\theta}]\|_p-2\kappa\varepsilon$. We can take e.g. $\kappa=2/3$, and it implies if $\|\hat{\theta}^{[\ell]}-\mathbb{E}[\hat{\theta}]\|_p<\varepsilon$ holds for ratio $2/3$ of all ℓ, we must have $\|\hat{\theta}_K-\mathbb{E}[\hat{\theta}]\|_p<4\varepsilon$. Using the Hoeffding bound, we have

$$\begin{aligned}\Pr\left[\|\hat{\theta}_K-\mathbb{E}[\hat{\theta}]\|_p\geq4\varepsilon\right]&\leq\Pr\left[\sum_{\ell=1}^{K}\mathbb{I}(\|\hat{\theta}^{[\ell]}-\mathbb{E}[\hat{\theta}]\|_p\geq\varepsilon)\geq\frac{K}{3}\right]\\&\leq\exp\left(-2K\left(\frac{1}{3}-\Pr[\|\hat{\theta}-\mathbb{E}[\hat{\theta}]\|_p\geq\varepsilon]\right)^2\right)\\&<\exp\left(-2K\left(\frac{1}{3}-\delta\right)^2\right).\end{aligned}\tag{D.15}$$

When $p<\infty$, using Jensen's inequality,

$$\begin{aligned}\sum_{i=1}^{m}\mathbb{E}[|\hat{\theta}_{K,i}-\theta_i|^2]^{p/2}&\leq\sum_{i=1}^{m}\mathbb{E}[|\hat{\theta}_{K,i}-\theta_i|^p]=\mathbb{E}[\|\hat{\theta}_K-\theta\|_p^p]\\&\leq\Pr[\|\hat{\theta}_K-\theta\|_p<4\varepsilon](4\varepsilon)^p+\Pr[\|\hat{\theta}_K-\theta\|_p\geq4\varepsilon]mu^p\\&<\exp\Big(-2K\Big(\frac{1}{3}-\delta\Big)^2\Big)mu^p+\Big(1-\exp\Big(-2K\Big(\frac{1}{3}-\delta\Big)^2\Big)\Big)(4\varepsilon)^p.\end{aligned}\tag{D.16}$$

When $p=\infty$,

$$\max_i\mathbb{E}[|\hat{\theta}_{K,i}-\theta_i|^2]<\exp\Big(-2K\Big(\frac{1}{3}-\delta\Big)^2\Big)u^2+\Big(1-\exp\Big(-2K\Big(\frac{1}{3}-\delta\Big)^2\Big)\Big)(4\varepsilon)^2.\tag{D.17}$$

In both cases, when $K=\Omega\big(\log\big(m^{1/p}/\varepsilon\big)\big)$, we can have

$$\left(\sum_{i=1}^{m}\mathbb{E}[(\hat{\theta}_{K,i}-\theta_i)^2]^{p/2}\right)^{1/p}\leq4\sqrt{2}\varepsilon.\tag{D.18}$$

3. Given an unbiased, bounded estimator that achieves a p-norm estimation error with probability $>2/3$ which takes N copies, and use the estimator $K=\Theta(\log\big(m^{1/p}/\varepsilon\big))$ times to calculate a

Geometric Median estimator, we obtain an unbiased estimator on $N_1 = N \times K$ copies satisfying Eq. (D.3) with $\xi = 4\sqrt{2}\varepsilon$. The CR bound implies

$$N = \frac{N_1}{K} = \Omega\left(\sup_{\rho_0\in\mathcal{S}^\circ} \frac{\left\|\operatorname{diag}\left((I(\rho_0,M)^{-1})_{AA}\right)^{1/2}\right\|_p}{\varepsilon^2\log\left(m^{1/p}/\varepsilon\right)}\right). \tag{D.19}$$

Here we generalize the above discussion to adaptive single-copy measurements. Consider an adaptive measurement on N copies of quantum states defined by $M^{(1,N)} = \{M^{(1,N)}_{x_1,\ldots,x_N}\} \in \mathcal{M}_{1,N}$ where

$$M^{(1,N)}_{x_1,\ldots,x_N} = M^{(1)}_{x_1} \otimes M^{(2)}_{x_1,x_2} \cdots \otimes M^{(N)}_{x_1,x_2,\ldots,x_N}, \tag{D.20}$$

where $M^{(r)}_{x_1,\ldots,x_{r-1},x_r}$ is the POVM operator on the r-th copy that depends on all previous outcomes $(x_1,\ldots,x_{r-1})$ with measurement outcome x_r. The superscript in $M^{(1,N)}$ means M acts on states from the 1st to the Nth copy. Using the CR bound, we have for any unbiased estimator $\hat{\theta}^{(N_1)}$,

$$V(\rho_{\theta,\varphi},(M^{(1,N)})^{\otimes K},\hat{\theta}^{(N_1)}) \succeq \frac{1}{N_1}(NI(\rho_{\theta,\varphi}^{\otimes N},M^{(1,N)})^{-1})_{AA}, \tag{D.21}$$

which is a generalization of Eq. (D.6). Similarly, Eq. (D.10) generalizes to

$$\xi \geq \frac{1}{\sqrt{N_1}}\sup_{\rho_0\in\mathcal{S}^\circ}\operatorname{tr}\left(\operatorname{diag}\left((NI(\rho_0^{\otimes N},M^{(1,N)})^{-1})_{AA}\right)^{p/2}\right)^{1/p}. \tag{D.22}$$

The lower bound needs to apply to all adaptive POVMs, i.e.,

$$\xi \geq \frac{1}{\sqrt{N_1}}\inf_{M^{(1,N)}\in\mathcal{M}_{1,N}}\sup_{\rho_0\in\mathcal{S}^\circ}\operatorname{tr}\left(\operatorname{diag}\left((NI(\rho_0^{\otimes N},M^{(1,N)})^{-1})_{AA}\right)^{p/2}\right)^{1/p}. \tag{D.23}$$

To prove Eq. (D.1) holds, we only need to show

$$\begin{aligned} v^{(N)} &:= \inf_{M^{(1,N)}\in\mathcal{M}_{1,N}}\sup_{\rho_0\in\mathcal{S}^\circ}\left\|\operatorname{diag}\left((NI(\rho_0^{\otimes N},M^{(1,N)})^{-1})_{AA}\right)^{1/2}\right\|_p \\ &= \inf_{M\in\mathcal{M}}\sup_{\rho_0\in\mathcal{S}^\circ}\left\|\operatorname{diag}\left((I(\rho_0,M)^{-1})_{AA}\right)^{1/2}\right\|_p =: v^{(1)}. \end{aligned} \tag{D.24}$$

First, we note that $v^{(N)} \leq v^{(1)}$ by definition. On the other hand, $v^{(N)} \geq v^{(1)}$ holds because the FIM of adaptive measurements can be written as

$$\begin{aligned} I(\rho_0^{\otimes r},M^{(1,r)}) &= I(\{p_{x_1,x_2,\ldots,x_{r-k}}\}) + \sum_{(x_1,x_2,\ldots,x_{r-k})} p_{x_1,x_2,\ldots,x_{r-k}} I(\rho_0^{\otimes k}, M^{(r-k+1,r)}_{x_1,x_2,\ldots,x_{r-k}}), \\ &= I(\rho_0^{\otimes r-k},M^{(1,r-k)}) + \sum_{(x_1,x_2,\ldots,x_{r-k})} p_{x_1,x_2,\ldots,x_{r-k}} I(\rho_0^{\otimes k}, M^{(r-k+1,r)}_{x_1,x_2,\ldots,x_{r-k}}), \end{aligned} \tag{D.25}$$

where $p_{x_1,x_2,\ldots,x_{r-k}}$ is the probability of obtaining measurement outcomes $x_1,x_2,\ldots,x_{r-k}$ and $M^{(r-k+1,r)}_{x_1,x_2,\ldots,x_{r-k}}$ is the POVM acting on states from the $r-k+1$ to the rth copy that depends on all previous measurement outcomes. Applying this decomposition trick multiple times, we have

$$\begin{aligned}I(\rho_0^{\otimes N}, M^{(1,N)}) &= I(\rho_0, M^{(1)}) + \sum_{x_1} p_{x_1} I(\rho_0, M^{(2)}_{x_1}) + \cdots \\ &\qquad + \sum_{x_1,\cdots,x_{N-1}} p_{x_1,\cdots,x_{N-1}} I(\rho_0, M^{(N)}_{x_1,\cdots,x_{N-1}}) \\ &= N I(\rho_0, \tilde{M}),\end{aligned} \tag{D.26}$$

where $\tilde{M}$ includes the POVM operators $\left\{\frac{1}{N}M^{(1)}_{x_1}, \frac{1}{N}p_{x_1}M^{(2)}_{x_1,x_2}, \ldots, \frac{1}{N}p_{x_1,\ldots,x_{N-1}}M^{(N)}_{x_1,\ldots,x_N}\right\}$ and is a single-copy measurement. This implies $v^{(N)} \geq v^{(1)}$, proving the theorem.

■

Theorem 20 (Lower bound for oblivious estimation using bounded and unbiased estimators) *Using the adaptive measurement strategy with single-copy measurements, the sample complexity of ρ required to obtain a bounded, unbiased estimator of θ_α for all α satisfying $\|\alpha\|_q \leq 1$ in Problem 2' is*

$$N = \Omega\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{\varepsilon^2 \log(1/\varepsilon)}\right), \tag{D.27}$$

for any $\varepsilon > 0$ and $p \in [1, \infty]$. Here "bounded" means the value of estimator $\hat{\theta}_\alpha$ is always away from the true value by a constant for any α.

Proof Similar to the proof of Theorem 19, the proof consists of three steps.

1. First, we will show that for a fixed non-adaptive single-copy measurement strategy $M^{\otimes N_1}$, we need a sample complexity of

$$N_1 = \Omega\left(\sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \leq 1} \frac{\alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha}{\xi^2}\right) \tag{D.28}$$

to construct an unbiased estimator $\hat{\theta}_\alpha$ that achieves a MSE smaller than ξ for all α satisfying $\|\alpha\|_q \leq$ 1, i.e.,

$$\sum_x (\hat{\theta}(x)_\alpha - \theta_\alpha)^2 \mathrm{tr}(\rho_\theta M_x) < \xi^2. \tag{D.29}$$

Given N_1 copies of parameterized quantum state

$$\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a \in A}\theta_a Q_a + \frac{1}{d}\sum_{b \in B}\varphi_b T_b, \tag{D.30}$$

where ρ_0 is known, $\{\theta_a\}_{a \in A}$ are to be estimated and $\{\varphi_b\}_{b \in B}$ are unknown (i.e. nuisance parameters), the CR bound states for any unbiased estimator $(\hat{\theta}^{(N_1)}, \varphi^{(N_1)})$,

$$V(\rho_{\theta,\varphi}, M^{\otimes N_1}, (\hat{\theta}^{(N_1)}, \varphi^{(N_1)})) \succeq \frac{1}{N_1} I(\rho_{\theta,\varphi}, M)^{-1}, \tag{D.31}$$

where $V(\rho_{\theta,\varphi}, M^{\otimes N_1}, (\hat{\theta}^{(N_1)}, \varphi^{(N_1)}))$ and $I(\rho_{\theta,\varphi}, M)$ are the MSEM and the FIM with respect to both parameters $\{\theta_a\}_{a\in A}$ and $\{\varphi_b\}_{b\in B}$. We can take the upper left blocks of the matrices that only involve entries in A, which gives

$$V(\rho_{\theta,\varphi}, M^{\otimes N_1}, \hat{\theta}^{(N_1)}) \succeq \frac{1}{N_1}(I(\rho_{\theta,\varphi}, M)^{-1})_{AA}, \tag{D.32}$$

Furthermore, we have

$$\alpha^\top V(\rho_{\theta,\varphi}, M^{\otimes N_1}, \hat{\theta}^{(N_1)})\alpha = \sum_x (\hat{\theta}(x)_\alpha - \theta_\alpha)^2 \mathrm{tr}(\rho_\theta M_x) \tag{D.33}$$

That implies when Eq. (D.29) holds for all α,

$$\xi \geq \frac{1}{\sqrt{N_1}} \max_{\|\alpha\|_q \leq 1} \alpha^\top (I(\rho_{\theta,\varphi}, M)^{-1})_{AA}\alpha. \tag{D.34}$$

Since we would like the above to hold for arbitrary (θ, φ) such that $\rho_{\theta,\varphi}$ is well defined. Then

$$\begin{aligned} \xi &\geq \frac{1}{\sqrt{N_1}} \sup_{(\theta,\varphi)\in\mathscr{D}(\rho_0)} \max_{\|\alpha\|_q \leq 1} \alpha^\top (I(\rho_{\theta,\varphi}, M)^{-1})_{AA}\alpha \\ &= \frac{1}{\sqrt{N_1}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\|\alpha\|_q \leq 1} \alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha, \end{aligned} \tag{D.35}$$

where the last equality holds because choosing a specific value of (θ, φ) is equivalent to replacing the original ρ_0 with $\rho_{\theta,\varphi}$ and then setting its value to be zero.

2. Next, we show for $p \in [1, \infty]$, if unbiased, bounded estimators $\hat{\theta}_\alpha$ satisfies for all α

$$|\hat{\theta}_\alpha - \theta_\alpha| < \varepsilon, \tag{D.36}$$

with probability $> 1 - \delta$, and

$$|\hat{\theta}_\alpha - \theta_\alpha| \leq u, \forall\alpha, \text{ s.t. } \|\alpha\|_q \leq 1, \tag{D.37}$$

for all measurement outcomes and some constant u, as required by Problem 2', then we can construct another unbiased estimator $\hat{\theta}_{\alpha,\mathrm{Med}}$ that satisfies Eq. (D.29) with $\xi = 2\varepsilon$ and overhead $K = O\big(\log(1/\varepsilon)\big)$. Without loss of generality, we assume $\delta < 1/3$.

Taking K independent samples of $\hat{\theta}_\alpha$, we define

$$\hat{\theta}_{\alpha,\mathrm{Med}} := \mathrm{Median}(\hat{\theta}^{[1]}, \ldots, \hat{\theta}^{[K]}), \tag{D.38}$$

where $\hat{\theta}^{[\ell]}$ are the ℓ-th sample of $\hat{\theta}_\alpha$. One property of the median estimator is if $|\hat{\theta}^{[\ell]} - \mathbb{E}[\hat{\theta}]| < \varepsilon$ holds for ratio $\kappa > 1/2$ of all ℓ, we must have $|\hat{\theta}_{\alpha,\mathrm{Med}} - \mathbb{E}[\hat{\theta}]| < \varepsilon$. Using the Hoeffding bound, we have

$$\Pr\left[|\hat{\theta}_{\alpha,\mathrm{Med}} - \mathbb{E}[\hat{\theta}]| \geq \varepsilon\right] \leq \Pr\left[\sum_{\ell=1}^{K} \mathbb{I}(|\hat{\theta}^{[\ell]} - \mathbb{E}[\hat{\theta}]| \geq \varepsilon) \geq \frac{K}{2}\right] < \exp\left(-2K\left(\frac{1}{2} - \delta\right)^2\right). \tag{D.39}$$

Then

$$\sum_x (\hat{\theta}(x)_\alpha - \theta_\alpha)^2 \mathrm{tr}(\rho_\theta M_x) < \exp\left(-2K\left(\frac{1}{2}-\delta\right)^2\right)u^2 + \left(1-\exp\left(-2K\left(\frac{1}{2}-\delta\right)^2\right)\right)\varepsilon^2. \tag{D.40}$$

In both cases, when $K = \Omega\big(\log(1/\varepsilon)\big)$, we can achieve

$$\sum_x (\hat{\theta}(x)_\alpha - \theta_\alpha)^2 \mathrm{tr}(\rho_\theta M_x) \le (2\varepsilon)^2. \tag{D.41}$$

3. Given an unbiased, bounded $\hat{\theta}_\alpha$ estimator that achieves ε error with probability $> 2/3$ which takes N copies, and use the estimator $K = \Theta(\log(1/\varepsilon))$ times to calculate a Median estimator, we obtain an unbiased estimator on $N_1 = N \times K$ copies satisfying Eq. (D.29) with $\xi = 2\varepsilon$. The CR bound implies

$$N = \frac{N_1}{K} = \Omega\left(\frac{\sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \le 1} \alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha}{\varepsilon^2 \log(1/\varepsilon)}\right). \tag{D.42}$$

Here we generalize the above discussion to adaptive single-copy measurements. Consider an adaptive measurement on N copies of quantum states defined by $M^{(1,N)} = \{M^{(1,N)}_{x_1,\ldots,x_N}\} \in \mathcal{M}_{1,N}$ where

$$M^{(1,N)}_{x_1,\ldots,x_N} = M^{(1)}_{x_1} \otimes M^{(2)}_{x_1,x_2} \cdots \otimes M^{(N)}_{x_1,x_2,\ldots,x_N}, \tag{D.43}$$

where $M^{(r)}_{x_1,\ldots,x_{r-1},x_r}$ is the POVM operator on the r-th copy that depends on all previous outcomes $(x_1,\ldots,x_{r-1})$ with measurement outcome x_r. The superscript in $M^{(1,N)}$ means M acts on states from the 1st to the Nth copy. Using the CR bound, we have for any unbiased estimator $\hat{\theta}^{(N_1)}$,

$$V(\rho_{\theta,\varphi}, (M^{(1,N)})^{\otimes K}, \hat{\theta}^{(N_1)}) \succeq \frac{1}{N_1}(NI(\rho_{\theta,\varphi}^{\otimes N}, M^{(1,N)})^{-1})_{AA}, \tag{D.44}$$

which is a generalization of Eq. (D.32). Then

$$\xi \ge \frac{1}{\sqrt{N_1}} \inf_{M^{(1,N)} \in \mathcal{M}_{1,N}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \le 1} \alpha^\top (NI(\rho_0^{\otimes N}, M^{(1,N)})^{-1})_{AA}\alpha. \tag{D.45}$$

To prove Eq. (D.1) holds, we only need to show

$$\begin{aligned} v^{(N)} &:= \inf_{M^{(1,N)} \in \mathcal{M}_{1,N}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \le 1} \alpha^\top (NI(\rho_0^{\otimes N}, M^{(1,N)})^{-1})_{AA}\alpha \\ &= \inf_{M \in \mathcal{M}} \sup_{\rho_0 \in \mathcal{S}^\circ} \max_{\|\alpha\|_q \le 1} \alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha =: v^{(1)}. \end{aligned} \tag{D.46}$$

First, we note that $v^{(N)} \le v^{(1)}$ by definition. On the other hand, $v^{(N)} \ge v^{(1)}$ holds because the FIM of adaptive measurements can be written as

$$\begin{aligned} I(\rho_0^{\otimes r}, M^{(1,r)}) &= I(\{p_{x_1,x_2,\ldots,x_{r-k}}\}) + \sum_{(x_1,x_2,\ldots,x_{r-k})} p_{x_1,x_2,\ldots,x_{r-k}} I(\rho_0^{\otimes k}, M^{(r-k+1,r)}_{x_1,x_2,\ldots,x_{r-k}}), \\ &= I(\rho_0^{\otimes r-k}, M^{(1,r-k)}) + \sum_{(x_1,x_2,\ldots,x_{r-k})} p_{x_1,x_2,\ldots,x_{r-k}} I(\rho_0^{\otimes k}, M^{(r-k+1,r)}_{x_1,x_2,\ldots,x_{r-k}}), \end{aligned} \tag{D.47}$$

where $p_{x_1,x_2,\ldots,x_{r-k}}$ is the probability of obtaining measurement outcomes $x_1, x_2, \ldots, x_{r-k}$ and $M^{(r-k+1,r)}_{x_1,x_2,\ldots,x_{r-k}}$ is the POVM acting on states from the $r-k+1$ to the rth copy that depends on all previous measurement outcomes. Applying this decomposition trick multiple times, we have

$$\begin{aligned} I(\rho_0^{\otimes N}, M^{(1,N)}) &= I(\rho_0, M^{(1)}) + \sum_{x_1} p_{x_1} I(\rho_0, M^{(2)}_{x_1}) + \cdots \\ &\qquad + \sum_{x_1,\cdots,x_{N-1}} p_{x_1,\cdots,x_{N-1}} I(\rho_0, M^{(N)}_{x_1,\cdots,x_{N-1}}) \\ &= N I(\rho_0, \tilde{M}), \end{aligned} \tag{D.48}$$

where $\tilde{M}$ includes the POVM operators $\left\{\frac{1}{N} M^{(1)}_{x_1}, \frac{1}{N} p_{x_1} M^{(2)}_{x_1,x_2}, \ldots, \frac{1}{N} p_{x_1,\ldots,x_{N-1}} M^{(N)}_{x_1,\ldots,x_N}\right\}$ and is a single-copy measurement. This implies $v^{(N)} \geq v^{(1)}$, proving the theorem.

■

D.2. Few-copy measurements

We then consider protocols using $(c \geq 1)$-copy measurements in $\mathcal{M}_c$. Similar to the distinguishing task in Theorem 18, we show that if we only care about ε^2 term, which is the case when ε is below a certain threshold, then the lower bound for these unbiased estimators for Problem 2 can again only achieve at most an $O(c)$ reduction from single-copy unbiased estimators. Formally, we have the following theorem

Theorem 21 (Lower bound for estimation and oblivious estimation using bounded, unbiased estimators, and c-copy measurements) *Using (possibly adaptive) measurement strategy with c-copy measurements, the sample complexity of ρ required to obtain a bounded, unbiased estimator of θ in Problem 2 (or a bounded, unbiased estimator of θ_α for all α satisfying $\|\alpha\|_q \leq 1$ in Problem 2')* *is*

$$N = \Omega\left(\frac{\Gamma_p(\{O_i\}_{i=1}^m)}{c\varepsilon^2 \log\left(m^{1/p}/\varepsilon\right)}\right), \tag{D.49}$$

for any $\varepsilon > 0$ and $p \in [2, \infty]$, or

$$N = \Omega\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{c\varepsilon^2 \log(1/\varepsilon)}\right), \tag{D.50}$$

for any $\varepsilon > 0$ and $p \in [1, \infty]$, respectively, where $I(\rho_0, M) = I(\rho_{\theta,\varphi}, M)|_{\theta=\varphi=0}$, $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A} \theta_a Q_a + \frac{1}{d}\sum_{b\in B} \varphi_b T_b$.

Proof First, we notice that following Theorem 19, to prove Eq. (D.49), we only need to prove

$$\inf_{M\in\mathcal{M}_c} \sup_{\rho_0\in\mathcal{S}^\circ} \left\|\mathrm{diag}\left((I(\rho_0, M)^{-1})_{AA}\right)^{1/2}\right\|_p^2 \geq \frac{1}{c^2} \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \left\|\mathrm{diag}\left((I(\rho_0, M)^{-1})_{AA}\right)^{1/2}\right\|_p^2. \tag{D.51}$$

The proof follows a similar argument with Theorem 18. Recall that

$$\chi^2_M(\rho_{\theta,\varphi}^{\otimes c}\|\rho_0^{\otimes c}) \le \left(\sum_{\varnothing\neq S\subseteq[c]} \sqrt{\sum_s \frac{\mathrm{tr}\left(M_s\left((\theta,\varphi)\cdot(\boldsymbol{Q},\boldsymbol{T})\right)^{(S)}\rho_0^{([c]\setminus S)}\right)^2}{d^2\,\mathrm{tr}(M_s\rho_0^{\otimes c})}}\right)^2, \tag{D.52}$$

for any c-copy measurement $M\in\mathcal{M}_c$. Taking the lowest order Taylor expansion (as in Eq. (B.25)) and using Eq. (C.101), we have

$$I(\rho_0^{\otimes c}, M)^{1/2} \preceq \sum_{i=1}^{c} I(\rho_0, G^{[i]})^{1/2}, \tag{D.53}$$

where $G^{[i]}$ is a single-copy POVM defined by

$$G_s^{[i]} := \mathrm{tr}_{[c]\setminus\{i\}}\left((\mathbb{I}^{(i)}\otimes\rho_0^{([c]\setminus\{i\})})M_s\right). \tag{D.54}$$

Note that

$$\left(\sqrt{I}-\sqrt{J}\right)^2 \succeq 0 \quad\Rightarrow\quad \sqrt{pI+(1-p)J} \succeq p\sqrt{I}+(1-p)\sqrt{J}, \forall 0\le p\le 1, \tag{D.55}$$

i.e. $\sqrt{(\cdot)}$ is operator concave. Then we have

$$\frac{1}{c}I(\rho_0^{\otimes c}, M)^{1/2} \preceq \frac{1}{c}\sum_{i=1}^{c} I(\rho_0, G^{[i]})^{1/2} \preceq \left(\frac{1}{c}\sum_{i=1}^{c} I(\rho_0, G^{[i]})\right)^{1/2} = I(\rho_0, G)^{1/2}. \tag{D.56}$$

where G is a single-copy POVM with measurement outcomes (i,s) such that $G_{(i,s)} := \frac{1}{c}G_s^{[i]}$. Eq. (D.51) is then proven using the above inequality, which means for any $M\in\mathcal{M}_c$ there exists some $G\in\mathcal{M}$ such that $I(\rho_0^{\otimes c}, M)\preceq c^2 I(\rho_0, G)$. Eq. (D.50) can be proven similarly. ■

Appendix E. Optimal estimator for high-precision shadow tomography

Here we construct an estimator that performs optimally for both Problem 2 and Problem 2' when the target precision is sufficiently small. We will first show the estimator that saturates the CR bound (up to a constant factor) for states in a neighborhood $\mathcal{N}(\rho_0)\subseteq\mathcal{S}$ of some specific state ρ_0. Then we introduce a tomography procedure that pre-determines ρ_0 such that our state $\rho\in\mathcal{N}(\rho_0)$. Finally, we analyze the performance of the estimator in terms of the metrics in Problem 2 and Problem 2' and calculate the thresholds of precision below which the sample complexity is optimal.

E.1. Optimal unbiased estimator within local regions

Although the CR bound is in general not necessarily saturable with finite sample complexity, for our specific shadow tomography problem here, we show the optimal locally unbiased estimator also performs optimally as a globally unbiased estimator within local regions up to a factor of two. The linearity of the probability distribution in (θ,φ) is the key property we use here.

Lemma 22 (Optimal estimation within local regions) *Given $\rho_0 \in \mathcal{S}^\circ$, consider quantum states parameterized as $\rho_{\theta,\varphi} = \rho_0 + \frac{1}{d}\sum_{a\in A}\theta_a Q_a + \frac{1}{d}\sum_{b\in B}\varphi_b T_b$, where $(\theta,\varphi) \in \mathscr{N}(\rho_0) = \mathscr{D}(\rho_0) \cap -\mathscr{D}(\rho_0)$. For any fixed POVM M, there exists an unbiased estimator $\hat{\theta}$ of θ such that the corresponding MSEM*

$$V(\rho_{\theta,\varphi}, M, \hat{\theta}) \preceq 2(I(\rho_0, M)^{-1})_{AA}, \tag{E.1}$$

where $I(\rho_0, M) = I(\rho_{\theta,\varphi}, M)|_{\theta=\varphi=0}$.

Proof First, we note that $\mathscr{N}(\rho_0)$ is the maximal region that satisfies the follow two properties.

- It is a closed, connected region.
- $(\theta,\varphi) \in \mathscr{N}(\rho_0) \Leftrightarrow (-\theta,-\varphi) \in \mathscr{N}(\rho_0)$.

The first property holds because $\mathscr{D}(\rho_0)$ is a closed region, and all $\theta \in \mathscr{D}(\rho_0)$ is connected to ρ_0, and the second property directly follows from the definition of $\mathscr{N}(\rho_0)$.

Let $p_{\theta,\varphi}(x) = \mathrm{tr}(\rho_{\theta,\varphi} M_x)$ for some POVM $M = \{M_x\}_x$. We define the following estimator which is the optimal locally unbiased estimator at $\theta,\varphi = 0$:

$$\hat{\theta}_a^{\mathrm{opt},\rho_0}(y) = \sum_{c,x}(I(\rho_0, M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}\delta_{xy}, \tag{E.2}$$

where we use index c to represent all parameters in $A \cup B$, and y is the measurement outcome, which is exactly the optimal locally estimator (Eq. (B.28)) at $\theta,\varphi = 0$.

First, we show it is an unbiased estimator within $\mathscr{N}(\rho_0)$. Let

$$f(\theta,\varphi)_a := \mathbb{E}_{\theta,\varphi}[\hat{\theta}_a^{\mathrm{opt},\rho_0}(x)] = \sum_{c,x}(I(\rho_0, M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0} p_{\theta,\varphi}(x), \tag{E.3}$$

where the expectation is taken for measurement on state $\rho_{\theta,\varphi}$. Then

$$f(0,0)_a = \sum_{c,x}(I(\rho_0, M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)} p_{\theta,\varphi}(x)\bigg|_{\theta,\varphi=0} = 0. \tag{E.4}$$

For any $(\theta,\varphi) \in \mathscr{N}^\circ(\rho_0)$, where we use $\mathscr{N}^\circ$ to denote the interior of $\mathscr{N}$,

$$\begin{aligned}
\partial_{c'} f(\theta,\varphi)_a &= \sum_{c,x}(I(\rho_0, M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}\partial_{c'} p_{\theta,\varphi}(x) \\
&= \sum_{c,x}(I(\rho_0, M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}\partial_{c'} p_{\theta,\varphi}(x)\big|_{\theta,\varphi=0} \\
&= \sum_{c'}(I(\rho_0, M)^{-1})_{ac} I(\rho_0, M)_{cc'} = \delta_{ac}, \quad \forall(\theta,\varphi) \in \mathscr{N}(\rho_0),
\end{aligned} \tag{E.5}$$

where we use the fact that $p_{\theta,\varphi}(x)$ is a linear function for all parameters, and thus

$$\partial_{c'} p_{\theta,\varphi}(x) = \partial_{c'} p_{\theta,\varphi}(x)\big|_{\theta,\varphi=0}. \tag{E.6}$$

The above implies the unbiasedness of the estimator, i.e.

$$f(\theta,\varphi)=\theta,\quad \forall(\theta,\varphi)\in\mathscr{N}(\rho_0), \tag{E.7}$$

because $\mathscr{N}(\rho_0)$ is closed and connected.

Next, we calculate the covariance matrix (i.e. the MSEM) of the estimator.

$$g_{aa'}(\theta,\varphi)=\mathbb{E}_{\theta,\varphi}[(\hat{\theta}^{\text{opt},\rho_0}_a-\theta_a)(\hat{\theta}^{\text{opt},\rho_0}_{a'}-\theta_{a'})]=\mathbb{E}_{\theta,\varphi}[\hat{\theta}^{\text{opt},\rho_0}_a\hat{\theta}^{\text{opt},\rho_0}_{a'}]-\theta_a\theta_{a'}. \tag{E.8}$$

Specifically,

$$\begin{aligned} g_{aa}(\theta,\varphi)+\theta_a^2&=\mathbb{E}_{\theta,\varphi}[(\hat{\theta}^{\text{opt},\rho_0}_a)^2]\\ &=\sum_{x,c,c'}(I(\rho_0,M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}(I(\rho_0,M)^{-1})_{ac'}\frac{\partial_{c'} p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}p_{\theta,\varphi}(x), \end{aligned} \tag{E.9}$$

$$\begin{aligned} g_{aa}(-\theta,-\varphi)+\theta_a^2&=\mathbb{E}_{-\theta,-\varphi}[(\hat{\theta}^{\text{opt},\rho_0}_a)^2]\\ &=\sum_{x,c,c'}(I(\rho_0,M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}(I(\rho_0,M)^{-1})_{ac'}\frac{\partial_{c'} p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}p_{-\theta,-\varphi}(x), \end{aligned} \tag{E.10}$$

$$\begin{aligned} &g_{aa}(\theta,\varphi)+g_{aa}(-\theta,-\varphi)+2\theta_a^2\\ &=2\sum_{x,c,c'}(I(\rho_0,M)^{-1})_{ac}\frac{\partial_c p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}(I(\rho_0,M)^{-1})_{ac'}\frac{\partial_{c'} p_{\theta,\varphi}(x)}{p_{\theta,\varphi}(x)}\bigg|_{\theta,\varphi=0}p_{0,0}(x)\\ &=2\sum_{c,c'}(I(\rho_0,M)^{-1})_{ac}I(\rho_0,M)_{cc'}(I(\rho_0,M)^{-1})_{ac'}=2(I(\rho_0,M)^{-1})_{aa}, \end{aligned} \tag{E.11}$$

where we use the fact that $p_{\theta,\varphi}(x)+p_{-\theta,-\varphi}(x)=2p_{0,0}(x)$. Analogously, we can show for any real m-dimensional vector $(v_a)_a$,

$$\sum_{aa'}v_av_{a'}(g_{aa'}(\theta,\varphi)+g_{aa'}(-\theta,-\varphi)+2\theta_a\theta_{a'})=\sum_{aa'}v_av_{a'}2(I(\rho_0,M)^{-1})_{aa'}, \tag{E.12}$$

which implies

$$g(\theta,\varphi)\preceq g(\theta,\varphi)+g(-\theta,-\varphi)\preceq 2(I(\rho_0,M)^{-1})_{AA}. \tag{E.13}$$

It means the MSEM of the estimator $\hat{\theta}^{\text{opt},\rho_0}$ is upper bounded by $2(I(\rho_0,M)^{-1})_{AA}$, proving the lemma. A crucial assumption we use above implicitly is $\rho_{-\theta,-\varphi}$ is well-defined, which is guaranteed because $(\theta,\varphi)\in\mathscr{N}(\rho_0)\Leftrightarrow(-\theta,-\varphi)\in\mathscr{N}(\rho_0)$. ■

E.2. Finding ρ_0 via state tomography

Here we discuss given an unknown state ρ, how to find ρ_0 such that $\rho\in\mathscr{N}(\rho_0)$ so that the above estimator applies. Here we abuse the notation a bit and say $\rho\in\mathscr{N}(\rho_0)$ if and only if $\rho=\rho_{\theta,\varphi}$ for some $(\theta,\varphi)\in\mathscr{N}(\rho_0)$. Finding ρ_0 is in general a difficult task, especially when ρ is singular. However, we can without loss of generality consider only states within a restricted set of states:

$$\mathcal{S}_{1/2}=\Big\{\rho\Big|\rho=\frac{1}{2}\Big(\sigma+\frac{\mathbb{I}}{d}\Big),\ \text{for some density matrix }\sigma\Big\}. \tag{E.14}$$

We can always assume $\rho \in \mathcal{S}_{1/2}$, because if not, we can apply the following quantum channel on the unknown state ρ

$$\rho \mapsto \frac{1}{2}\rho + \frac{\mathbb{I}}{2d} \tag{E.15}$$

whose output state belongs to $\mathcal{S}_{1/2}$ and then perform the estimation on the output state. The channel maps the expectation values $\mathrm{tr}(O_i\rho) \mapsto \frac{1}{2}\mathrm{tr}(O_i\rho)$, which induces at most a constant factor in the estimation precision and the sample complexity bounds. In this case, it is sufficient to find an estimator of ρ that is within $1/(4d)$ of its operator norm.

Lemma 23 *For any state $\rho \in \mathcal{S}_{1/2}$, if $\|\rho_0 - \rho\|_\infty \leq 1/(4d)$, $\rho \in \mathscr{N}(\rho_0)$.*

Proof We first note that $\rho \in \mathscr{N}(\rho_0)$ if and only if

$$\rho_0 - (\rho - \rho_0) \succeq 0 \quad \Leftrightarrow \quad \rho - (\rho - \rho_0) - (\rho - \rho_0) \succeq 0. \tag{E.16}$$

It holds when $\|\rho_0 - \rho\|_\infty \leq 1/(4d)$ because

$$\rho - (\rho - \rho_0) - (\rho - \rho_0) \succeq \rho - 2\|\rho - \rho_0\|_\infty \mathbb{I} \succeq \rho - \mathbb{I}/(2d) \succeq 0. \tag{E.17}$$

■

The above implies that any algorithm that produces an estimate of ρ such that its ∞-norm distance to ρ is at most $1/(4d)$ will be sufficient to serve as the first step to determine ρ_0, prior to applying the unbiased estimator $\hat{\theta}^{\mathrm{opt},\rho_0}$ in the second step. To achieve the target accuracy in tomography with single-copy measurements, one can apply the Haar random measurement $\{d\,|v\rangle\,\langle v|\,\mathrm{d}v\}$ and use the estimator

$$\hat{\rho}_0(N_0, u) = \frac{1}{N_0}\sum_{i=1}^{N_0}((d+1)\,|u_i\rangle\,\langle u_i| - \mathbb{I}), \tag{E.18}$$

where $u = (u_1, \ldots, u_{N_0})$ are the outcomes from measuring ρ. With probability $1-\delta$ (Kueng et al., 2017; Guţă et al., 2020; Chen and Cotler, 2025),

$$\|\hat{\rho}_0(N_0, u) - \rho\|_\infty \leq \text{Constant} \times \max\left\{\frac{d + \log(1/\delta)}{N}, \sqrt{\frac{d + \log(1/\delta)}{N}}\right\}. \tag{E.19}$$

In particular, $\hat{\rho}_0(N_0, u)$ must be a well-defined density matrix when $\rho \in \mathcal{S}_{1/2}$ and $\|\hat{\rho}_0(N_0, u) - \rho\|_\infty \leq 1/4d$. Therefore, $N_0 = O(d^3)$ is sufficient to achieve the target precision with high probability and find a local region $\mathscr{N}(\rho_0)$ to apply the optimal local estimator.

E.3. Conversion from MSEM to p-norm error

We showed in Appendix E.1, the optimal locally unbiased estimator $\hat{\theta}^{\mathrm{opt},\rho_0}$ performs optimally in estimating θ in $\rho_{\theta,\varphi}$ within $\mathscr{N}(\rho_0)$ in the sense that it achieves, up to a factor of two, the optimal MSEM given by the CR bound. Our goal is to learn observables with p-norm error. Below we show, using the (coordinate-wise) median-of-means estimator, we can, with probability at least $1-\delta$, obtain a bounded p-norm error using the estimator of bounded p-average RMSE with an overhead of $O(\log(m/\delta))$.

Lemma 24 *Fix the POVM M. For any unbiased estimator $\hat{\theta}$ that achieves*

$$\left\|\operatorname{diag}(V(\rho_{\theta,\varphi}, M, \hat{\theta}))^{1/2}\right\|_p \leq \xi, \tag{E.20}$$

there is another estimator $\hat{\theta}_{MoM}$ that achieves p-norm error ε with probability at least $1-\delta$ that uses

$$O\left(\log\left(\frac{m}{\delta}\right)\right) \times O\left(\frac{\xi^2}{\varepsilon^2}\right) \tag{E.21}$$

samples of $\hat{\theta}$.

Proof Given an unbiased estimator $\hat{\theta}$ for $\theta \in \mathbb{R}^m$ satisfying the p-average RMSE bound, i.e.

$$\left\|\operatorname{diag}(V(\rho_{\theta,\varphi}, M, \hat{\theta}))^{1/2}\right\|_p = \left(\sum_{j=1}^{m} \sigma_j^p\right)^{1/p} \leq \xi, \tag{E.22}$$

where $\sigma_j^2 := \operatorname{Var}[\hat{\theta}_j] = V(\rho_{\theta,\varphi}, M, \hat{\theta})_{jj}$ is the variance of estimating θ_j. Our goal is to construct an estimator $\hat{\theta}_{\mathrm{MoM}}$ such that with probability at least $1-\delta$:

$$\left\|\hat{\theta}_{\mathrm{MoM}} - \theta\right\|_p \leq \varepsilon \tag{E.23}$$

using a sample complexity overhead that is poly-logarithmic in m and $1/\delta$.

We start with the definition of the coordinate-wise median-of-means estimator:

1. *Sampling:* Collect $N_1 = K \times B$ independent samples of $\hat{\theta}$. Divide them into K batches, each of size B.
2. *Batch Averaging:* For each batch $\ell \in \{1, \ldots, K\}$, compute the empirical mean:

$$\hat{\theta}_B^{[\ell]} := \frac{1}{B}\sum_{i=1}^{B} \hat{\theta}^{[\ell,i]}, \tag{E.24}$$

where $\hat{\theta}^{[\ell,i]}$ is the i-th sample of $\hat{\theta}$ in the ℓ-th batch.

3. *Coordinate-wise Median:* For each coordinate $j \in \{1, \ldots, m\}$, compute the median of the batch means:

$$\hat{\theta}_{\mathrm{MoM},j} = \operatorname{Median}\left(\hat{\theta}_{B,j}^{[1]}, \ldots, \hat{\theta}_{B,j}^{[K]}\right). \tag{E.25}$$

We now try to analyze the performance of the estimator.

1. *Bounding probability for a single batch (Chebyshev).* Fix a coordinate j, the variance of the batch mean $\hat{\theta}_{B,j}^{[\ell]}$ is σ_j^2/B. By Chebyshev's inequality, for any ℓ:

$$\Pr\left[\left|\hat{\theta}_{B,j}^{[\ell]} - \theta_j\right| > 2\frac{\sigma_j}{\sqrt{B}}\right] \leq \frac{\operatorname{Var}[\hat{\theta}_{B,j}^{[\ell]}]}{(2\sigma_j/\sqrt{B})^2} = \frac{\sigma_j^2/B}{4\sigma_j^2/B} = \frac{1}{4}. \tag{E.26}$$

Let us define the "bad" event for the ℓ-th batch on coordinate j as $\mathcal{E}_{\ell,j} := \left\{|\hat{\theta}_{B,j}^{[\ell]} - \theta_j| > 2\frac{\sigma_j}{\sqrt{B}}\right\}$. We have established that $\Pr[\mathcal{E}_{\ell,j}] \leq 1/4$.

2. *Bounding probability for the Median (Chernoff/Hoeffding).* For the median $\hat{\theta}_{\text{MoM},j}$ to deviate from θ_j by more than $2\sigma_j/\sqrt{B}$, more than half of the batches must satisfy the bad event $\mathcal{E}_{\ell,j}$. Let $S_j = \sum_{\ell=1}^{K} \mathbb{I}[\mathcal{E}_{\ell,j}]$, where $\mathbb{I}[\cdot]$ is the indicator function. The expected number of bad batches is $\mathbb{E}[S_j] \leq K/4$. The failure condition for the median is $S_j \geq K/2$. Using Hoeffding's inequality:

$$\begin{aligned} \Pr\left[\left|\hat{\theta}_{\text{MoM},j} - \theta_j\right| > 2\frac{\sigma_j}{\sqrt{B}}\right] &\leq \Pr\left[S_j \geq \frac{K}{2}\right] \\ &\leq \exp\left(-2K\left(\frac{1}{2} - \frac{1}{4}\right)^2\right) = \exp\left(-\frac{K}{8}\right). \end{aligned} \tag{E.27}$$

3. *Union Bound over Coordinates.* We require all coordinates to satisfy their respective bounds simultaneously to preserve the sum-structure of the p-norm. Apply the union bound over all m coordinates. Let $\mathcal{F}$ be the event that *any* coordinate j fails (i.e., $|\hat{\theta}_{\text{MoM},j} - \theta_j| > 2\frac{\sigma_j}{\sqrt{B}}$ for some j).

$$\Pr[\mathcal{F}] \leq \sum_{j=1}^{m} \exp\left(-\frac{K}{8}\right) = m \exp\left(-\frac{K}{8}\right). \tag{E.28}$$

To ensure this failure probability is at most δ, we set:

$$m \exp\left(-\frac{K}{8}\right) \leq \delta \implies K \geq 8 \ln\left(\frac{m}{\delta}\right). \tag{E.29}$$

4. *Bounding the p-norm error.* Conditioned on the success event $\mathcal{F}^c$ (the complement of failure, which occurs with probability $\geq 1 - \delta$), we have that for all $j \in \{1, \dots, m\}$:

$$\left|\hat{\theta}_{\text{MoM},j} - \theta_j\right| \leq \frac{2\sigma_j}{\sqrt{B}}. \tag{E.30}$$

Now, we compute the p-norm of the error vector:

$$\left\|\hat{\theta}_{\text{MoM}} - \theta\right\|_p = \left(\sum_{j=1}^{m} \left|\hat{\theta}_{\text{MoM},j} - \theta_j\right|^p\right)^{1/p} \tag{E.31}$$

Substituting the coordinate-wise bounds:

$$\left\|\hat{\theta}_{\text{MoM}} - \theta\right\|_p \leq \left(\sum_{j=1}^{m} \left(\frac{2\sigma_j}{\sqrt{B}}\right)^p\right)^{1/p} = \frac{2}{\sqrt{B}} \left(\sum_{j=1}^{m} \sigma_j^p\right)^{1/p}. \tag{E.32}$$

Using the initial assumption that the p-average RMSE is bounded by ξ:

$$\left\|\hat{\theta}_{\text{MoM}} - \theta\right\|_p \leq \frac{2\xi}{\sqrt{B}}. \tag{E.33}$$

To achieve a target p-norm error of ε, we set $\frac{2\xi}{\sqrt{B}} \leq \varepsilon$, which implies $B \geq \frac{4\xi^2}{\varepsilon^2}$. The total number of samples required is:

$$N_1 = K \times B = O\left(\log\left(\frac{m}{\delta}\right)\right) \times O\left(\frac{\xi^2}{\varepsilon^2}\right). \tag{E.34}$$

Thus, using the coordinate-wise median-of-means estimator, we obtain a bounded p-norm error using the estimator of bounded p-average RMSE with an overhead of $O(\log(m/\delta))$. ■

E.4. Algorithm, sample complexity, and threshold

Here we describe the algorithm that combines the three steps introduced above, analyze the corresponding sample complexity and derive the threshold below which the sample complexity matches our lower bound (up to logarithmic overhead). We consider the case of shadow tomography (Problem 1, equivalent to Problem 2) and the case of oblivious single-observable estimation (Problem 1', equivalent to Problem 2') separately below.

E.4.1. SHADOW ESTIMATION WITH p-NORM ERROR

Theorem 25 (Upper bound for estimation with p-norm error) *For any $\varepsilon > 0$ and $p \in [1,\infty]$ there exists an algorithm that uses*

$$\begin{aligned} N &= O(d^3) + O\left(\frac{\log(m)}{\varepsilon^2} \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \left\| \operatorname{diag}\left((I(\rho_0,M)^{-1})_{AA}\right)^{1/2}\right\|_p^2\right) \\ &= O(d^3) + O\left(\frac{\log(m)\Gamma_p(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right) \end{aligned} \tag{E.35}$$

copies of ρ and single-copy measurements to solve Problem 1 (or equivalently, Problem 2). In particular, when

$$\begin{aligned} \varepsilon \le \overline{\eta} &= \sqrt{\frac{\log(m)\Gamma_p(\{O_i\}_{i=1}^m)}{d^3}}, \\ N &= O\left(\frac{\log(m)\Gamma_p(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right). \end{aligned} \tag{E.36}$$

Proof We first pick a POVM $M^\diamond \in \mathcal{M}$ that achieves

$$\sup_{\rho_0\in\mathcal{S}^\circ} \left\|\operatorname{diag}(I(\rho_0,M^\diamond)^{-1}_{AA})^{1/2}\right\|_p^2 \le 2 \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \left\|\operatorname{diag}(I(\rho_0,M)^{-1}_{AA})^{1/2}\right\|_p^2 = 2\Gamma_p(\{O_i\}_{i=1}^m). \tag{E.37}$$

Then we consider the following algorithm in Algorithm 1:

Here picking

$$\begin{aligned} N_0 = O(d^2(d+\log(1/\delta))), \quad N_1 = O\left(\frac{\log(m)}{\varepsilon^2}\Gamma_p(\{O_i\}_{i=1}^m)\right), \\ K = O\left(\log\left(\frac{m}{\delta}\right)\right), \quad B = O\left(\frac{1}{\varepsilon^2}\Gamma_p(\{O_i\}_{i=1}^m)\right), \end{aligned} \tag{E.38}$$

is sufficient to guarantee the desired performance of our algorithm. To see this, we first notice by Lemma 23, $\rho \in \mathscr{N}(\rho_0)$ with high probability. Then using Lemma 22, we have

$$V(\rho_{\theta,\varphi}, M^\diamond, \hat{\theta}^{\mathrm{opt},\rho_0}) \preceq 2(I(\rho_0,M^\diamond)^{-1})_{AA}, \tag{E.39}$$

and

$$\left\|\operatorname{diag}(V(\rho_{\theta,\varphi}, M^\diamond, \hat{\theta}^{\mathrm{opt},\rho_0})^{1/2})\right\|_p^2 \le 2\left\|\operatorname{diag}((I(\rho_0,M^\diamond)^{-1})_{AA})^{1/2})\right\|_p^2 \le 4\Gamma_p(\{O_i\}_{i=1}^m). \tag{E.40}$$

```
Algorithm 1: Shadow tomography with bounded $p$-norm error
Input: Observables $\{O_i\}_{i=1}^m$, $p \in [1,\infty]$, $\delta, \varepsilon > 0$, $N = N_0 + N_1$ copies of a $d$-dimensional state $\rho$
Output: Estimators $\hat{o} = (\hat{o}_i, \ldots, \hat{o}_m)$
Goal: With probability at least $1-\delta$, $(\sum_{i=1}^m |\mathrm{tr}(O_i\rho) - \hat{o}_i|^p)^{1/p} < \varepsilon$
1. Mix all $\rho$'s evenly with the maximally mixed state: $\rho \mapsto \frac{1}{2}\rho + \frac{\mathbb{I}}{2d}$
2. Apply the Haar random measurement on each of the $N_0$ copies of $\rho$ and obtain $\rho_0 = \hat{\rho}_0(N_0, u)$
   as a coarse estimate $\rho$, where $u = (u_1, \ldots, u_{N_0})$ are the measurement outcomes, such that
   $\|\rho_0 - \rho\|_\infty \leq 1/4d$ with high probability
3. Apply POVM $M^\diamond$ on each of the $N_1$ copies of $\rho$ and obtain $N_1$ unbiased estimates of $\theta$ using the
   optimal unbiased estimator $\hat{\theta}^{\mathrm{opt},\rho_0}$
4. Divide the $N_1$ estimates into $K$ groups each with $B$ elements and calculate the coordinate-wise
   median-of-means estimator $\hat{\theta}_{\mathrm{MoM}}$
Return: $\hat{o}_i = 2((\hat{\theta}_{\mathrm{MoM}})_i + \mathrm{tr}(O_i\rho_0))$
```

Using Lemma 24, it then follows that a coordinate-wise median-of-means estimator with $K = O(\log(m/\delta))$ groups of $B = O(\Gamma_p(\{O_i\}_{i=1}^m)/\varepsilon^2)$ elements is sufficient for our purpose. ■

Note that from (Chiribella et al., 2007) (see also the proof of Lemma 17), here we can choose $M^\diamond$ as a POVM with at most d^2 measurement outcomes.

E.4.2. OBLIVIOUS SINGLE-OBSERVABLE ESTIMATION

Theorem 26 (Upper bound for oblivious single-observable estimation) *For any $\varepsilon > 0$ and $p \in [1,\infty]$ there exists an algorithm that uses*

$$\begin{aligned} N &= O(d^3) + O\left(\frac{1}{\varepsilon^2} \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\substack{\alpha\in\mathbb{R}^m\\ \|\alpha\|_q\leq 1}} \alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha\right) \\ &= O(d^3) + O\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right) \end{aligned} \tag{E.41}$$

copies of ρ and single-copy measurements to solve Problem 1' (or equivalently, Problem 2'). When

$$\begin{aligned} \varepsilon \leq \overline{\eta}^{\mathrm{ob}} &= \sqrt{\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{d^3}}, \\ N &= O\left(\frac{\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m)}{\varepsilon^2}\right). \end{aligned} \tag{E.42}$$

Proof We first pick a POVM $M^\diamond \in \mathcal{M}$ that achieves

$$\begin{aligned} &\sup_{\rho_0\in\mathcal{S}^\circ} \max_{\substack{\alpha\in\mathbb{R}^m\\ \|\alpha\|_q\leq 1}} \alpha^\top (I(\rho_0, M^\diamond)^{-1})_{AA}\alpha \\ &\qquad \leq 2 \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \max_{\substack{\alpha\in\mathbb{R}^m\\ \|\alpha\|_q\leq 1}} \alpha^\top (I(\rho_0, M)^{-1})_{AA}\alpha = 2\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m). \end{aligned} \tag{E.43}$$

```
Algorithm 2: Oblivious estimation of $O_\alpha = \sum_i \alpha_i O_i$ for $\|\alpha\|_q \le 1$
Input: Observables $\{O_i\}_{i=1}^m$, $p \in [1,\infty]$, $\delta, \varepsilon > 0$, copies of a $d$-dimensional state $\rho$
Output: Estimator $\hat{o}_\alpha$
Goal: With probability at least $1-\delta$, $|\operatorname{tr}(\rho O_\alpha) - \hat{o}_\alpha| < \varepsilon$
1. Mix all $\rho$'s evenly with the maximally mixed state: $\rho \mapsto \frac{1}{2}\rho + \frac{\mathbb{I}}{2d}$
2. Apply the Haar random measurement on each of the $N_0$ copies of $\rho$ and obtain $\rho_0 = \hat{\rho}_0(N_0, u)$
   as a coarse estimate $\rho$, where $u = (u_1, \ldots, u_{N_0})$ are the measurement outcomes, such that
   $\|\rho_0 - \rho\|_\infty \le 1/4d$ with high probability
3. Apply POVM $M^\diamond$ on each of the $N_1$ copies of $\rho$ and obtain $N_1$ unbiased estimates of $\theta$ using
   the optimal unbiased estimator $\hat{\theta}^{\mathrm{opt},\rho_0}$
4. Reveal $\alpha \in \mathbb{R}^m$, which satisfies $\|\alpha\|_q \le 1$
5. Divide the $N_1$ samples of estimates $\hat{\theta}_\alpha := \sum_{i=1}^m \alpha_i \hat{\theta}_i^{\mathrm{opt},\rho_0}$ into $K$ groups each with $B$ elements
   and calculate the median-of-means estimator $\hat{\theta}_{\mathrm{MoM},\alpha}$
Return: $\hat{o}_\alpha = 2(\hat{\theta}_{\mathrm{MoM},\alpha} + \operatorname{tr}(\rho_0 O_\alpha))$
```

Then we consider the following algorithm in Algorithm 2:

Here picking

$$N_0 = O(d^2(d + \log(1/\delta))), \quad N_1 = O\left(\frac{\log(1/\delta)}{\varepsilon^2}\Gamma_p(\{O_i\}_{i=1}^m)\right), \\ K = O(\log(1/\delta)), \quad B = O\left(\frac{1}{\varepsilon^2}\Gamma_p(\{O_i\}_{i=1}^m)\right), \tag{E.44}$$

is sufficient to guarantee the desired performance of our algorithm. To see this, we first notice by Lemma 23, $\rho \in \mathscr{N}(\rho_0)$ with high probability. Then using Lemma 22, we have

$$V(\rho_{\theta,\varphi}, M^\diamond, \hat{\theta}^{\mathrm{opt},\rho_0}) \preceq 2(I(\rho_0, M^\diamond)^{-1})_{AA}, \tag{E.45}$$

and

$$\alpha^\top V(\rho_{\theta,\varphi}, M^\diamond, \hat{\theta}^{\mathrm{opt},\rho_0})\alpha \le 2\alpha^\top (I(\rho_0, M^\diamond)^{-1})_{AA}\alpha \le 4\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m). \tag{E.46}$$

Furthermore, let $\hat{\theta}_\alpha = \sum_{i=1}^m \alpha_i (\hat{\theta}^{\mathrm{opt},\rho_0})_i$, we have

$$\operatorname{Var}[\hat{\theta}_\alpha] = \alpha^\top V(\rho_{\theta,\varphi}, M^\diamond, \hat{\theta}^{\mathrm{opt},\rho_0})\alpha \le 4\Gamma_p^{\mathrm{ob}}(\{O_i\}_{i=1}^m). \tag{E.47}$$

We can consider a simpler variant of Lemma 24, where only a single parameter $\theta_\alpha = \sum_{i=1}^{\alpha} \alpha_i \theta_i$ is to be estimated, which implies whenever $\operatorname{Var}[\hat{\theta}_\alpha] \le \xi^2$, $K = O(\log(1/\delta))$ and $B = O(\xi^2/\varepsilon^2)$ samples can guarantee the Median-of-Mean estimator $\hat{\theta}_{\mathrm{MoM},\alpha}$ has an additive error within $\varepsilon/2$ with probability at least $1-\delta$, which achieves our desired precision. ■

Note that from (Chiribella et al., 2007) (see also the proof of Lemma 17), it is always possible to choose $M^\diamond$ as a POVM with at most d^2 measurement outcomes. Also notice that although in the proofs above we pick $M^\diamond$ to be the near-optimal estimators that satisfy Eq. (E.37) and Eq. (E.43), respectively. In fact, the proofs still hold if we pick $M^\diamond$ adaptively where we remove the state optimization $\sup_{\rho_0 \in \mathcal{S}^\circ}$ from Eq. (E.37) and Eq. (E.43) and fix ρ_0 to be the coarse state estimate from the first step. This could potentially reduce the classical computational cost to identify $M^\diamond$.

E.5. Simple relation between Γ_p and Γ_p^{ob}

Due to the tightness of our bounds for both the oblivious estimation task and the shadow estimation task, we already know that

$$\Gamma_p^{\text{ob}}(\{O_i\}_{i=1}^m) = O(\Gamma_p(\{O_i\}_{i=1}^m)) \tag{E.48}$$

for all $p \in [1, \infty]$, which means Γ_p^{ob} is no larger than Γ_p up to constant. Here we show the constant is simply 1.

Lemma 27 *For all $p \in [1, \infty]$, $\Gamma_p^{\text{ob}}(\{O_i\}_{i=1}^m) \le \Gamma_p(\{O_i\}_{i=1}^m)$. The equality holds when $p = \infty$.*

Proof To prove the inequality, we only need to show for any matrix $R \succeq 0$,

$$\max_{\|\alpha\|_q \le 1} \alpha^\top R \alpha \le \left\|\operatorname{diag}(R)^{1/2}\right\|_p^2. \tag{E.49}$$

Let v_i denote the i-th column of $\sqrt{R}$. Then for any α,

$$\sqrt{\alpha^\top R \alpha} = \|\sqrt{R}\alpha\|_2 = \left\|\sum_{i=1}^m \alpha_i v_i\right\|_2 \le \sum_{i=1}^m |\alpha_i| \|v_i\|_2. \tag{E.50}$$

Moreover, $\|v_i\|_2^2 = R_{ii}$. We have

$$\sqrt{\alpha^\top R \alpha} \le \sum_{i=1}^m |\alpha_i| \sqrt{R_{ii}}. \tag{E.51}$$

For any $\|\alpha\|_q \le 1$, by Hölder's inequality,

$$\sqrt{\alpha^\top R \alpha} = \sum_{i=1}^m |\alpha_i| \sqrt{R_{ii}} \le \left(\sum_{i=1}^m \sqrt{R_{ii}}^p\right)^{1/p} = \left\|\operatorname{diag}(R)^{1/2}\right\|_p, \tag{E.52}$$

proving Eq. (E.49). Finally, we note that when $p = \infty$ and $q = 1$,

$$\max_{\|\alpha\|_1 \le 1} \alpha^\top R \alpha = \max_i R_{ii} = \left\|\operatorname{diag}(R)^{1/2}\right\|_\infty^2. \tag{E.53}$$

This is because $\alpha^\top R \alpha$ is a convex function in α, and the maximum can be taken at extreme points, i.e. when α has only one entry equal to one. ■

Appendix F. Example: Pauli estimation

We showcase our results with a concrete example of Pauli observable estimation. While the case of $p = \infty$ has been thoroughly investigated in (Chen et al., 2024b) with tight sample complexity bound obtained, our framework extends the analysis to all $p \in [1, \infty]$ in the high-precision regime. Furthermore, our bounds are tight up to log factors for all $p \ge 2$.

Theorem 28 (Oblivious learning of Pauli observables, single-copy measurements) *Let* $d = 2^n$ *and* $\{O_i := P_i\}_{i=1}^{d^2-1}$ *be all* n*-qubit traceless Pauli operators. Then,*

$$\Omega(d) \leq \Gamma_p^{\mathrm{ob}} \leq \begin{cases} O(d \log d)\,, & \text{if } p \in [2, \infty], \\ O\Big(d^{\frac{4}{p}-1} \log d\Big)\,, & \text{if } p \in [1, 2). \end{cases} \tag{F.1}$$

Furthermore, the following sample complexity bounds for the oblivious learning problems (Problem 1') hold:

$$\begin{aligned} &N = \Omega(\Gamma_p^{\mathrm{ob}}/\varepsilon^2) = \Omega(d/\varepsilon^2), && \text{when } \varepsilon < \frac{1}{6} d^{\frac{2}{p}-2}. \\ &N = O(\Gamma_p^{\mathrm{ob}}/\varepsilon^2) = \begin{cases} O\left(d \log d/\varepsilon^2\right), & \text{if } p \in [2, \infty], \\ O\left(d^{\frac{4}{p}-1} \log d/\varepsilon^2\right), & \text{if } p \in [1, 2), \end{cases} && \text{when } \varepsilon < c_0 d^{-1}. \end{aligned} \tag{F.2}$$

Here $c_0 > 0$ *is some absolute constant.*

Proof When $O_i = P_i$ for all $i \in [m]$ where $m = d^2 - 1$, it is easy to see that the dual operator basis is uniquely determined as $Q_i = P_i$ for all $i \in [m]$. Recall the dual representation of Γ_p^{ob} thanks to Lemma 13,

$$(\Gamma_p^{\mathrm{ob}})^{-1} = \sup_{M \in \mathcal{M}} \inf_{\rho_0 \in \mathcal{S}^\circ} \min_{\|\theta\|_p = 1} \theta^\top I(\rho_0, M) \theta. \tag{F.3}$$

Let us first derive an upper bound for this,

$$\begin{aligned} (\Gamma_p^{\mathrm{ob}})^{-1} &\leq \sup_{M \in \mathcal{M}} \min_{\|\theta\|_p = 1} \theta^\top I(\mathbb{I}/d, M) \theta \\ &\leq \sup_{M \in \mathcal{M}} \min_{i \in [m]} I(\mathbb{I}/d, M)_{ii} \\ &= \sup_{M \in \mathcal{M}} \min_{i \in [m]} \sum_x \frac{\mathrm{tr}^2(M_x P_i)}{d \mathrm{tr}(M_x)} \\ &\leq \sup_{M \in \mathcal{M}} \frac{1}{m} \sum_{i \in [m]} \sum_x \frac{\mathrm{tr}^2(M_x P_i)}{d \mathrm{tr}(M_x)} \\ &\leq \sup_{M \in \mathcal{M}} \frac{d}{d^2 - 1} \sum_x \frac{\mathrm{tr}(M_x^2)}{d \mathrm{tr}(M_x)} \\ &\leq \frac{d}{d^2 - 1} = O(d^{-1}). \end{aligned} \tag{F.4}$$

The first line fixes $\rho_0 = \mathbb{I}/d$. The second line restricts θ to one-hot vectors. The fifth line uses the twirling formula for Pauli operators: $\sum_{P \in \mathcal{P}_n} \mathrm{tr}^2(P M_x) = d \mathrm{tr}(M_x^2)$. The last line uses $\mathrm{tr}(M_x^2) \leq \mathrm{tr}^2(M_x)$ and the normalization condition of POVMs.

Next, we derive a lower bound. By fixing the M to be the Haar random measurement $\mathcal{M}_\mu := \{d \left|\psi\right\rangle\!\left\langle\psi\right|\}_{\psi \sim \mu}$, where μ is the Haar measure over d-dimensional pure states:

$$\begin{aligned} (\Gamma_p^{\mathrm{ob}})^{-1} &\geq \inf_{\rho_0 \in \mathcal{S}^\circ} \min_{\|\theta\|_p = 1} \theta^\top I(\rho_0, M_\mu) \theta \\ &= \inf_{\rho_0 \in \mathcal{S}^\circ} \min_{\|\theta\|_p = 1} \mathbb{E}_{\psi \in \mu} \frac{\left\langle\psi\right| \sum_i \theta_i P_i \left|\psi\right\rangle^2}{d \left\langle\psi\middle|\rho_0\middle|\psi\right\rangle}. \end{aligned} \tag{F.5}$$

For any $x > 0$ and $k \in \mathbb{N}_+$, we have the following concentration bound:

$$\Pr_{\psi\sim\mu}\left(\langle\psi|\rho_0|\psi\rangle \geq x\right) \leq \frac{\mathbb{E}_{\psi\sim\mu}\langle\psi|\rho_0|\psi\rangle^k}{x^k} = \frac{\operatorname{tr}\left(\rho_0^{\otimes k}\Pi_k\right)}{\binom{d+k-1}{k}x^k} \leq \frac{k!}{d^k x^k} \leq \frac{\sqrt{2\pi k}k^k e^{\frac{1}{12k}}}{e^k d^k x^k}. \tag{F.6}$$

The first line uses the k-th order Markov inequality. The second line uses the Haar integral formula, where Π_k is the projector onto the symmetric subspace. The third line uses that $\Pi_k \leq \mathbb{I}$ and simple algebra on the binomial coefficients. The last line uses a non-asymptotic Stirling's upper bound on $k!$ (Robbins, 1955). Now we take $x := (c\log d)/d$ and $k = \lceil c\log d\rceil$ for some constant $c \geq 4$. The above inequality becomes

$$\Pr_{\psi\sim\mu}\left(\langle\psi|\rho_0|\psi\rangle \geq \frac{c\log d}{d}\right) \leq C\frac{\sqrt{\log d}}{d^c}. \tag{F.7}$$

For some constant $C > 0$ that depends only on c. Now, fix any $\rho_0 \in \mathcal{S}^\circ$ and $\theta \in \mathbb{R}^m$. Call ψ good if $\langle\psi|\rho_0|\psi\rangle < (c\log d)/d$ and bad elsewise. We have:

$$\begin{aligned}
\mathbb{E}_{\psi\in\mu}\frac{\langle\psi|\sum_i\theta_i P_i|\psi\rangle^2}{d\,\langle\psi|\rho_0|\psi\rangle} &\geq \int_{\psi:\text{good}} \mathrm{d}\mu_\psi \frac{\langle\psi|\sum_i\theta_i P_i|\psi\rangle^2}{d\,\langle\psi|\rho_0|\psi\rangle} \\
&> \frac{1}{c\log d}\int_{\psi:\text{good}} \mathrm{d}\mu_\psi \langle\psi|\sum_i\theta_i P_i|\psi\rangle^2 \\
&= \frac{1}{c\log d}\left(\int \mathrm{d}\mu_\psi \langle\psi|\sum_i\theta_i P_i|\psi\rangle^2 - \int_{\psi:\text{bad}} \mathrm{d}\mu_\psi \langle\psi|\sum_i\theta_i P_i|\psi\rangle^2\right) \\
&\geq \frac{1}{c\log d}\left(\frac{d\|\theta\|_2^2}{d(d+1)} - C\frac{\sqrt{\log d}}{d^c}\|\theta\|_1^2\right) \\
&\geq \frac{\|\theta\|_2^2}{c\log d}\left(\frac{1}{d+1} - C\frac{\sqrt{\log d}}{d^{c-2}}\right) \\
&= \|\theta\|_2^2\,\Omega\left(\frac{1}{d\log d}\right).
\end{aligned} \tag{F.8}$$

The first line restricts the integral to good ψ's. The second line uses the definitions of good ψ's. For the fourth line, the first term uses the Haar integral formula for $k = 2$ and $\operatorname{tr}(P_i) = 0$ and $\operatorname{tr}(P_iP_j) = d\delta_{ij}$; the second term uses $\langle\psi|P_i|\psi\rangle \leq 1$. The fifth line uses $\|\theta\|_1^2 \leq m\|\theta\|_2^2$. The last line holds as long as we take, say, $c \geq 4$. Put this back to Eq. (F.5):

$$(\Gamma_p^{\text{ob}})^{-1} \geq \min_{\|\theta\|_p=1}\|\theta\|_2^2\,\Omega\left(\frac{1}{d\log d}\right) = \begin{cases} \Omega\left(\dfrac{1}{d\log d}\right), & \text{if } p \in [2,\infty], \\ \Omega\left(\dfrac{1}{d^{\frac{4}{p}-1}\log d}\right), & \text{if } p \in [1,2). \end{cases} \tag{F.9}$$

This concludes our proof for the bounds on Γ_p^{ob}.

Now we prove the claimed thresholds. The lower bound threshold immediately follows from Theorem 14:

$$\eta_p^{\text{ob}} = \frac{1}{6\,\|\|P_1\|_\infty,\cdots,\|P_m\|_\infty\|_q} = \frac{1}{6}(d^2-1)^{-\frac{1}{q}} \geq \frac{1}{6}d^{-\frac{2}{q}} = \frac{1}{6}d^{\frac{2}{p}-2}. \tag{F.10}$$

This means $\varepsilon \le$ R.H.S. is sufficient to guarantee $\varepsilon \le \eta_p^{\mathrm{ob}}$. For the upper bound threshold, combining Theorem 26 with our lower bound of $\Gamma_p^{\mathrm{ob}} = \Omega(d)$, we have

$$\overline{\eta}_p^{\mathrm{ob}} = \sqrt{\Gamma_p^{\mathrm{ob}}/d^3} = \Omega(d^{-1}). \tag{F.11}$$

Therefore, there exists an absolute constant $c_0 > 0$ such that when $\varepsilon < c_0 d^{-1}$ it is guaranteed that $\varepsilon < \overline{\eta}_p^{\mathrm{ob}}$. This completes the proof. ∎

Corollary 29 *For shadow estimation of complete Pauli observables,* $\Omega(d) \le \Gamma_2 \le O(d^3 \log d)$.

Proof The lower bound part follows from Lemma 27. The upper bound part can be seen as follows:

$$\begin{aligned}\Gamma_2 &= \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \operatorname{tr}(I_{\rho_0,M}^{-1}) \\ &\le m \inf_{M\in\mathcal{M}} \sup_{\rho_0\in\mathcal{S}^\circ} \left(\min_{\|\theta\|_2=1} \theta^\top I_{\rho_0,M}\theta \right)^{-1} \\ &= m\,\Gamma_2^{\mathrm{ob}} \\ &= O(d^3 \log d).\end{aligned} \tag{F.12}$$

The second line uses the variational expression for the minimal eigenvalue. ∎

We also consider Pauli observable estimation with few-copy measurements. We propose the following lower bound regarding Problem 2(2') and Problem 3(3').

Theorem 30 (Lower bound of Pauli observables for Problem 2(2') and Problem 3(3'), c-copy measurements) *Let $d = 2^n$ and $\{P_i\}_{i=1}^{d^2-1}$ be all n-qubit traceless Pauli operators. Using (possibly adaptive) measurement strategy with c-copy measurements, the sample complexity of ρ parameterized by $\{P_i\}_{i=1}^{d^2-1}$ required to obtain a bounded, unbiased estimator of θ in Problem 2 (or a bounded, unbiased estimator of θ_α for all α satisfying $\|\alpha\|_q \le 1$ in Problem 2') is*

$$N = \Omega\left(\frac{d}{c\varepsilon^2 \log(1/\varepsilon)}\right), \tag{F.13}$$

for any $\varepsilon > 0$ and $p \in [1,\infty]$.

Using (possibly adaptive) measurement strategy with c-copy measurements, consider the many-versus-one distinguishing tasks in Problem 3(3') with states ρ parameterized by $\{P_i\}_{i=1}^{d^2-1}$. The sample complexity required to solve this task is

$$N = \Omega\left(\frac{d}{c\varepsilon^2}\right), \tag{F.14}$$

for any

$$\varepsilon \le \begin{cases} \min\left\{ \frac{1}{6} d^{\frac{2}{p}-2}, \frac{c_1}{c d^{\frac{2}{p}-\frac{1}{2}}\sqrt{\log d}} \right\}, & \text{if } p \in [1,2), \\ \min\left\{ \frac{1}{6} d^{\frac{2}{p}-2}, \frac{c_1}{c d^{\frac{5}{2}-\frac{4}{p}}\sqrt{\log d}} \right\}, & \text{if } p \in [2,\infty], \end{cases} \tag{F.15}$$

where $c_1 > 0$ is an absolute constant.

Proof Eq. (F.13) and Eq. (F.14) can be obtained by combining Theorem 18 and Theorem 21 with Theorem 28. In the following, we focus on deriving Eq. (F.15). Note that with Theorem 28, we only need to compute (an upper bound on) $a_{\max}$ defined in Theorem 18. Recall that

$$a_{\max} := \max_{\rho_0 \in \mathcal{S}_{1/2}, \|\theta\|_p = 1} a_{\rho_0}(\theta), \tag{F.16}$$

$$a_{\rho_0}(\theta) = \theta^\top G^{(\rho_0)} \theta, \quad G_{ij}^{(\rho_0)} := \frac{\operatorname{tr}(P_i \rho_0^{-1} P_j)}{d^2}. \tag{F.17}$$

Given the (Hilbert–Schmidt) orthogonality

$$\operatorname{tr}(P_i P_j) = d\delta_{ij}, \tag{F.18}$$

For the ρ_0–weighted Gram matrix

$$G_{ij}^{(\rho_0)} := \frac{\operatorname{tr}\big(P_i \rho_0^{-1} P_j\big)}{d^2}, \tag{F.19}$$

its Rayleigh quotient on any $\theta \in \mathbb{R}^M$ is

$$\theta^\top G^{(\rho_0)} \theta = \frac{\operatorname{tr}\Big(\big(\sum_i \theta_i P_i\big) \rho_0^{-1} \big(\sum_j \theta_j P_j\big)\Big)}{d^2} = \frac{\operatorname{tr}(X \rho_0^{-1} X)}{d^2}, \quad X := \sum_i \theta_i P_i. \tag{F.20}$$

Using the operator bounds

$$\rho_0^{-1} \;\le\; \lambda_{\max}(\rho_0^{-1}) \mathbb{I} \tag{F.21}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue, we obtain

$$\theta^\top G^{(\rho_0)} \theta \le \frac{\lambda_{\max}(\rho_0^{-1}) \operatorname{tr}(X^2)}{d^2}. \tag{F.22}$$

Because $\{P_i\}$ are orthogonal with $\operatorname{tr}(P_i^2) = d$, we have

$$\operatorname{tr}(X^2) = \sum_{i,j} \theta_i \theta_j \operatorname{tr}(P_i P_j) = d\|\theta\|_2^2. \tag{F.23}$$

Therefore

$$\theta^\top G^{(\rho_0)} \theta \le \frac{\lambda_{\max}(\rho_0^{-1})}{d} \|\theta\|_2^2, \tag{F.24}$$

Using the standard ℓ_p–ℓ_2 extrema, we have

$$\max_{\|\theta\|_p = 1} \|\theta\|_2^2 \le \begin{cases} 1, & p \in [1, 2), \\ d^{2-4/p}, & p \in [2, \infty]. \end{cases} \tag{F.25}$$

Note that

$$\rho_0 = \frac{1}{2} \frac{\mathbb{I}}{d} + \frac{1}{2} \sigma, \tag{F.26}$$

for some state σ as $\rho_0 \in \mathcal{S}_{1/2}$. The spectrum of ρ_0 obeys

$$\lambda_{\min}(\rho_0) \;\geq\; \frac{1}{2} \cdot \frac{1}{d}, \tag{F.27}$$

hence

$$\lambda_{\max}(\rho_0^{-1}) \leq 2d, \tag{F.28}$$

and

$$a_{\max} \leq \begin{cases} 2, & p \in [1,2), \\ 2d^{2-4/p}, & p \in [2,\infty]. \end{cases} \tag{F.29}$$

Combining with Theorem 18, we immediately obtain Eq. (F.15). ■