ARTICLE   OPEN

A Bayesian analysis of classical shadows

Joseph M. Lukens [1 ✉], Kody J. H. Law[2] and Ryan S. Bennink[3]

The method of classical shadows proposed by Huang, Kueng, and Preskill heralds remarkable opportunities for quantum estimation with limited measurements. Yet its relationship to established quantum tomographic approaches, particularly those based on likelihood models, remains unclear. In this article, we investigate classical shadows through the lens of Bayesian mean estimation (BME). In direct tests on numerical data, BME is found to attain significantly lower error on average, but classical shadows prove remarkably more accurate in specific situations—such as high-fidelity ground truth states—which are improbable in a fully uniform Hilbert space. We then introduce an observable-oriented pseudo-likelihood that successfully emulates the dimension-independence and state-specific optimality of classical shadows, but within a Bayesian framework that ensures only physical states. Our research reveals how classical shadows effect important departures from conventional thinking in quantum state estimation, as well as the utility of Bayesian methods for uncovering and formalizing statistical assumptions.

INTRODUCTION

Measurement and characterization of quantum systems comprise a long-standing problem in quantum information science[1]. However, the exponential scaling of Hilbert space dimension with the number of qubits makes full characterization extremely challenging, inspiring a plethora of approaches designed to estimate properties of quantum states with as few measurements as possible, such as compressed sensing[2,3], adaptive tomography[4–6], matrix product state formulations[7], permutationally invariant tomography[8,9], and neural networks[10–12]. Very recently, a groundbreaking approach known as classical shadows was proposed and analyzed[13]. Building on and simplifying ideas from "shadow tomography"[14], the classical shadow was shown to provide accurate predictions of observables with a fixed number of measurements, including simulated examples for quantum systems in excess of 100 qubits[13]. Astonishingly simple, the classical shadow is formed by collecting the results of random measurements on a repeatedly prepared input state, and inverting them through an appropriate virtual quantum channel.

However, several features of the classical shadow remain enigmatic, including its highly nonphysical nature, optimality with respect to alternative cost functions, and relationship to more conventional likelihood-based tomographic techniques. One such method, Bayesian mean estimation (BME)[15], provides a conceptually straightforward path to estimate a quantum state given measured data, making use of prior knowledge, and providing meaningful error bars for any experimental conditions. BME appears particularly well-suited for contextualizing classical shadows, since it returns a principled estimate under any number of measurements (even zero), and is optimal in terms of minimizing average-squared error[16].

In this work, we directly compare the estimates of classical shadows and BME for identical simulated datasets. For particular observables with relatively improbable values from the perspective of BME, shadow is found to reach the ground truth with significantly fewer measurements. However, after properly reformulating the problem under test for consistency with the Bayesian prior, the situation reverses, with BME returning estimates

possessing lower error on average. In the latter portion of our investigation, we seek to construct a BME model emulating the key features of the classical shadow, but with positive semidefinite states as support. While complicated by the shadow's nonphysical nature, we ultimately propose an observable-oriented pseudo-likelihood that rates quantum states by their observable values with respect to those of shadow. Our pseudo-likelihood successfully mimics the dimension independence of shadow, with the advantage of delivering entirely physical estimates for any number of measurements.

RESULTS

Problem formulation

For our analysis, we invoke the setup of the original classical shadow proposal[13]. Consider a D-dimensional Hilbert space occupied by a ground truth quantum state ρ_g that can be repeatedly prepared. On each preparation m, ρ_g is subjected to a randomly chosen $D \times D$ unitary U_m and one measurement is performed in the computational basis, leaving result $|b_m\rangle$. Defining $|\psi_m\rangle = U_m^\dagger |b_m\rangle$, the classical snapshot associated with measurement m follows as $\mathcal{M}^{-1}(|\psi_m\rangle\langle\psi_m|)$, where $\mathcal{M}(\cdot)$ is the quantum channel defined by averaging over all possible unitaries and outcomes.

We assume the U_m are drawn from the set of $D \times D$ Haar-random unitaries, in which case $\mathcal{M}^{-1}(|\psi_m\rangle\langle\psi_m|) = (D + 1)|\psi_m\rangle\langle\psi_m| - I_D$, with I_D the $D \times D$ identity matrix[13]. (This channel holds for the more restricted class of random Cliffords as well[17,18].) Averaging over M measurements yields the shadow estimator

$$\rho_s = \frac{D+1}{M}\sum_{m=1}^{M}|\psi_m\rangle\langle\psi_m| - I_D. \tag{1}$$

(In what follows, the phrases "classical shadow," "shadow estimator," and simply "shadow" refer interchangeably to this estimator, as well as to the procedure more generally.) In this form, the simplicity of ρ_s is evident: it is merely a scaled and recentered average of all observed outcomes. Interestingly, though, ρ_s is in general not positive semidefinite; for $M < D$, ρ_s possesses at least

[1]Quantum Information Science Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [2]Department of Mathematics, University of Manchester, Manchester, UK. [3]Quantum Computational Science Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. ✉email: lukensjm@ornl.gov

$D - M$ eigenvalues equal to -1. Accordingly, in the targeted regime for classical shadows of $M \ll D$, ρ_s is highly nonphysical. Understanding the impact of the shadow estimator's negativity is a central goal of the present study. Finally, defining λ as the expectation of the observable Λ ($\lambda = \mathrm{Tr}\rho\Lambda$), the shadow estimate thereof follows as

$$\lambda^{(s)} = \mathrm{Tr}\,\rho_s\Lambda, \tag{2}$$

to be compared to the ground truth $\lambda^{(g)} = \mathrm{Tr}\,\rho_g\Lambda$.

As an aside, we note that ref. [13] employed an additional statistical technique, "median of means," to reduce the impact of outliers by partitioning the M outcomes into K subsets and taking the median as the estimate $\lambda^{(s)}$. In the interests of simplicity and ease of comparison, we focus on $K = 1$ in Eq. (1). We expect the benefits of selecting $K > 1$ would prove similar in both the shadow and Bayesian cases[19], but work on this is beyond the scope of the present investigation.

In the Bayesian paradigm, the same set of measurement outcomes $\mathcal{D} = \{|\psi_1\rangle, |\psi_2\rangle, ..., |\psi_M\rangle\}$ is related to a possible density matrix $\rho(\mathbf{x})$ via a likelihood consisting of the product of probabilities set by Born's rule:

$$L_{\mathcal{D}}(\mathbf{x}) = \prod_{m=1}^{M} \langle\psi_m|\rho(\mathbf{x})|\psi_m\rangle, \tag{3}$$

that is, $L_{\mathcal{D}}(\mathbf{x}) \propto \mathrm{Pr}(\mathcal{D}|\rho)$—the probability of receiving the dataset $\mathcal{D}$ given quantum state ρ. Some prior distribution $\pi_0(\mathbf{x})$ is also assumed, defined for parameters $\mathbf{x}$ such that $\rho(\mathbf{x})$ is always physical: trace-1, Hermitian, and positive semidefinite. Then the posterior describing the distribution of $\mathbf{x}$ given the observed data $\mathcal{D}$ ensues from Bayes' rule:

$$\pi(\mathbf{x}) = \frac{1}{\mathcal{Z}} L_{\mathcal{D}}(\mathbf{x})\pi_0(\mathbf{x}). \tag{4}$$

Note that the randomness of the chosen unitaries U_m does not enter the Bayesian model; only the outcomes $|\psi_m\rangle$ play a role. The selection of unitary U_m is independent of the (unknown) density matrix, i.e., $\mathrm{Pr}(U_m = U|\rho) = \mathrm{Pr}(U_m = U)$; thus any probabilities would cancel out through the normalization factor $\mathcal{Z}$ in Eq. (4). Intuitively, in the Bayesian view, the experimenter knows the unitaries exactly post-experiment, regardless of how they were chosen, so imposing uncertainty on them in the estimation process proves superfluous. Consequently, while the uncertainty of BME depends strongly on the variety of measurements chosen, the theory does not, a conspicuous departure from shadow where the distribution of U_m enters directly through the inverted quantum channel $\mathcal{M}^{-1}(\cdot)$.

Formally, the posterior distribution in Eq. (4) completes the Bayesian model. From this, one can estimate any function of $\rho(\mathbf{x})$, including uncertainties, since the full probability distribution $\pi(\mathbf{x})$ is available. The ability to quantify uncertainty in a single, consistent framework is an extremely attractive feature of Bayesian methods in general. In the present work, however, the goal is to compare Bayesian tomography with the classical shadow—which is itself a point estimator—so we focus on BME specifically, which for some observable Λ returns the point estimator defined as

$$\begin{aligned}\lambda^{(B)} &= \langle\mathrm{Tr}\,\rho\Lambda\rangle_\rho \\ &= \int d\mathbf{x}\,\pi(\mathbf{x})\,\mathrm{Tr}\,\rho(\mathbf{x})\Lambda \\ &= \mathrm{Tr}\left\{\left[\int d\mathbf{x}\,\pi(\mathbf{x})\rho(\mathbf{x})\right]\Lambda\right\} \\ &= \mathrm{Tr}\,\rho_B\Lambda,\end{aligned} \tag{5}$$

where the last two lines follow, respectively, from the linearity of the trace operation and defining the Bayesian mean $\rho_B = \int d\mathbf{x}\,\pi(\mathbf{x})\rho(\mathbf{x})$. This convenient simplification, in which the Bayesian mean of a quantity is simply its value at ρ_B, holds for linear functions of ρ, which includes all quantum observables and which we focus on in this article. In addition, selection of the mean as our particular Bayesian point estimator for comparison is motivated by the fact that it minimizes the mean-squared error (MSE) averaged over all possible states and outcomes. Specifically, $\lambda^{(B)}$ defined above is the function of $\mathcal{D}$ such that

$$\lambda^{(B)} = \underset{\lambda(\mathcal{D})}{\mathrm{argmin}} \int d\mathcal{D} \int d\mathbf{x}\,\pi(\mathbf{x}, \mathcal{D})[\lambda(\mathcal{D}) - \mathrm{Tr}\,\rho(\mathbf{x})\Lambda]^2, \tag{6}$$

with $\pi(\mathbf{x}, \mathcal{D})$ the joint distribution over data and parameters[16]. This optimality is nonasymptotic, holding for any number or collection of unitaries $\{U_1, U_2, ..., U_M\}$. Considering the widely different expressions for ρ_s [Eq. (1)] and ρ_B [Eq. (5)], we found it remarkable just how well ρ_s performed in ref. [13] in light of BME's optimality in Eq. (6); it was this feature which initially inspired us to develop a thorough comparison between shadow and BME. So while focusing on the Bayesian mean point estimator admittedly underutilizes the full depth of Bayesian methods, it provides a direct path for side-by-side comparison with classical shadows, along with a quantitative optimality condition.

In general, comparing the performance of estimators derived from classical (frequentist) statistics—like ρ_s—with those from Bayesian methods proves tricky business, since they view uncertainty in functionally different ways. Therefore, we adopt a pragmatic view which aligns with the interests of experimentalists: perform experiments, compute the associated shadow and BME estimators, and calculate their error with respect to actual values. While the final step is not always possible in practice, it is in numerical simulation, where the ground truth ρ_g is known exactly. Doing so enables us to illuminate the advantages and disadvantages of both approaches on equal footing. We employ the approach described in the "Methods" section for obtaining simulated datasets $\mathcal{D}$.

Picture 1: fixed ground truth

As our first benchmark, we compare the performance of ρ_s and ρ_B in estimating three rank-1 observables, of which fidelities and entanglement witnesses form an important and experimentally relevant subset. Specifically, we consider ground truth $\rho_g = |0\rangle\langle 0|$ and observables $\Lambda_n = |\phi_n\rangle\langle\phi_n|$ ($n = 0, 1, 2$) where

$$\begin{aligned}|\phi_0\rangle &= |0\rangle \\ |\phi_1\rangle &= \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2(D-1)}}\sum_{j=1}^{D-1}|j\rangle \\ |\phi_2\rangle &= |1\rangle.\end{aligned} \tag{7}$$

These possess ground truth values equally spaced within the physically allowed range for trace-1, rank-1 observables: $\lambda_0^{(g)} = 1$, $\lambda_1^{(g)} = \frac{1}{2}$, and $\lambda_2^{(g)} = 0$. The shadow estimator is readily obtained from Eq. (1), so we compute ρ_s for all $M \in \{1, 2, ..., 1000\}$, where M defines the set containing the first M measurements: $\mathcal{D} = \{|\psi_m\rangle;\ m = 1, 2, ..., M\}$.

On the other hand, ρ_B requires evaluation of the high-dimensional integral $\int d\mathbf{x}\,\pi(\mathbf{x})\rho(\mathbf{x})$. To that end, we summon Markov chain Monte Carlo (MCMC) methods, several of which have been explored in the context of quantum state estimation, including Metropolis–Hastings[15,20], Hamiltonian Monte Carlo[21], sequential Monte Carlo (SMC)[22], and slice sampling[23,24]; we select the preconditioned Crank–Nicolson (pCN) algorithm[25] applied to quantum state estimation in ref. [26]. Finally, because of our assumed pure state ground truth, we take as prior all pure states $\rho = |\psi\rangle\langle\psi|$ uniformly distributed on the complex D-dimensional unit hypersphere. Numerically, the parameters $\mathbf{x}$ reduce to a D-dimensional complex column vector, so we have $\pi_0(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^\dagger\mathbf{x}\right)$, $\rho(\mathbf{x}) = \frac{\mathbf{x}\mathbf{x}^\dagger}{|\mathbf{x}|^2}$, and $d\mathbf{x} = \prod_{l=1}^{D} d(\mathrm{Re}x_l)d(\mathrm{Im}x_l)$ with x_l denoting a single component of $\mathbf{x}$.
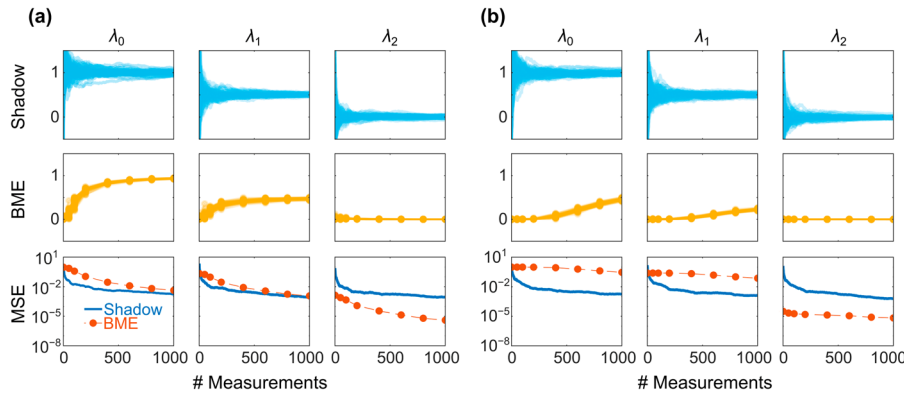
Fig. 1 Comparison of shadow and BME estimates of λ_n (Picture 1). a $D = 32$ case. **b** $D = 256$ case. Results from 50 trials for each dimension are plotted, assuming a fixed ground truth state $|0\rangle$.

The use of pure states is not central to the BME formalism whatsoever, but does permit us to simulate in higher dimensions than otherwise possible. With pure states only, our parameterization entails $2D$ real numbers, compared to $2D^2 + D$ for mixed states. As an example, for $D = 256$, the pure state prior, and likelihood of Eq. (3), each MCMC chain takes ~10 min to converge on our desktop computer, which for the 400 settings involved in Fig. 1b amounts to ~2.5 days. Based on previous studies[26], the mixed state version would therefore have been completely unfeasible at this dimension with our computational resources, likely taking weeks (or more) to complete. (Incorporating some of the methods suggested in ref. [26] in further research, such as embedding within SMC samplers and parallelization, should permit the extension to significantly larger D and mixed states.) With pure states, then, we can focus more directly on dimensional scaling and the statistics from many trials.

For each trial, we perform BME for eight collections of measurements $M \in \{1, 50, 100, 200, 400, 600, 800, 1000\}$. We keep $R = 2^{10}$ samples from each chain of length RT, where we select the thinning factor T empirically to obtain convergence. (See "Methods" for more details on the MCMC chains, including convergence metrics and marginal distributions.) Figure 1 plots the estimates for all 50 trials obtained by both shadow and BME for $D = 32$ (Fig. 1a) and $D = 256$ (Fig. 1b). A thinning value of $T = 2^9$ ($T = 2^{12}$) is used for $D = 32$ ($D = 256$). Each column corresponds to a particular expectation value λ_n; the bottom row shows the MSE with respect to the ground truth, averaged over all trials defined as $\left\langle |\lambda_n^{(\cdot)} - \lambda_n^{(g)}|^2 \right\rangle_{\text{trials}}$ with $\cdot = s$ for the shadow and $\cdot = B$ for BME. The classical shadows show wide variation for low M, including highly nonphysical estimates ($\lambda_n^{(s)} > 1$ or < 0), but they converge to ground truth values rapidly, with nearly identical rates for all observables and dimensions. This is confirmed quantitatively in the MSE curves that attain values of ~10^{-3} by $M = 1000$ for all cases.

The behavior proves vastly different for BME. While physical estimates are always returned, the number of measurements needed to reach the ground truth varies strongly both with observable λ_n and with dimension D. Intriguingly, shadow shows significantly lower MSE for λ_0 and λ_1, widening as D increases. On first glance, this presents a paradox: Eq. (6) implies that $\lambda_n^{(B)}$ should possess the lowest possible MSE for any n and M, and yet $\lambda_n^{(s)}$ convincingly surpasses it these cases. Yet this dilemma can be resolved by studying the prior $\pi_0(\mathbf{x})$. When the Bayesian model assigns equal a priori weights to all possible states—a sensible choice for an uninformative prior—this by implication makes observable values such as $\lambda_0^{(g)} = 1$ highly unlikely, since only one state in the domain attains this. On the other hand, expectations for any rank-1 projector Λ on the order of $\lambda \sim \frac{1}{D}$ are to be expected

initially since $\int d\mathbf{x}\, \pi_0(\mathbf{x}) \text{Tr}\, \rho(\mathbf{x})\Lambda = \frac{1}{D}$. This manifests itself in Fig. 1 in BME's much lower MSE for λ_2, whose ground truth value $\lambda_2^{(g)} = 0$ is much more probable. Thus, by running 50 repeated trials with the same ground truth $\rho_g = |0\rangle\langle 0|$, the situation over which we average does not accurately reflect the uninformative prior; the conditions for BME optimality are not met.

Picture 2: random ground truth

To accurately reflect uninformative prior knowledge, we therefore must prepare random ground truth states in our simulations. To do so, we leverage the equivalence between (i) randomly prepared input states with a fixed observable—the situation of interest—and (ii) random selection of an observable for a fixed input. Consider the expectation of observable Λ, where the quantum state is rotated by some random unitary U:

$$\text{Tr}\left[(U\rho U^{\dagger})\Lambda\right] = \text{Tr}\left[\rho(U^{\dagger}\Lambda U)\right]. \tag{8}$$

Thus, one can emulate the effect of a randomized state by randomly rotating the observable and evaluating it on a fixed state. Practically speaking, we are free to employ the same simulated datasets and estimators ρ_s and ρ_B above, but select at random a different projector $\Lambda = |\phi\rangle\langle\phi|$ for each trial. This is equivalent to performing all trials with a random ground truth, but a fixed observable. We call this randomized evaluation "Picture 2" to distinguish it from the fixed ground truth case above (Picture 1).

Results appear in Fig. 2 for (a) $D = 32$ and (b) $D = 256$. The first column plots the ground truth value $\lambda^{(g)}$ for each trial, the next three columns plot the shadow and BME estimates for increasing numbers of measurements, and the final column presents the MSE with respect to the ground truth. Now BME returns much more accurate estimates than shadow on average, and the paradox regarding Bayesian optimality is solved: the Bayesian mean gives the lowest MSE as long as the prior accurately reflects the true uncertainty of the system under test. Accordingly, this BME study clarifies an underlying assumption in selecting observables in Picture 1: being able to "guess" an observable with such high overlap to the ground truth suggests that one is not really operating under the neutrality implied by a uniform prior; an informative prior would more accurately reflect the situation.

This observation brings to light an interesting question of motivation in a given quantum experiment. In the sense of ensuring that any estimate is adequately justified by the data, the idea of "baking in" a prior favoring some subset of quantum states is undesirable. And yet, in many situations the researcher does have strong beliefs—or at least hopefulness—about the state being prepared, and wants to verify this by computing an observable, such as fidelity, where it is desired that $\lambda^{(g)} \sim 1$. In this case, one wishes to validate such high values quickly with few
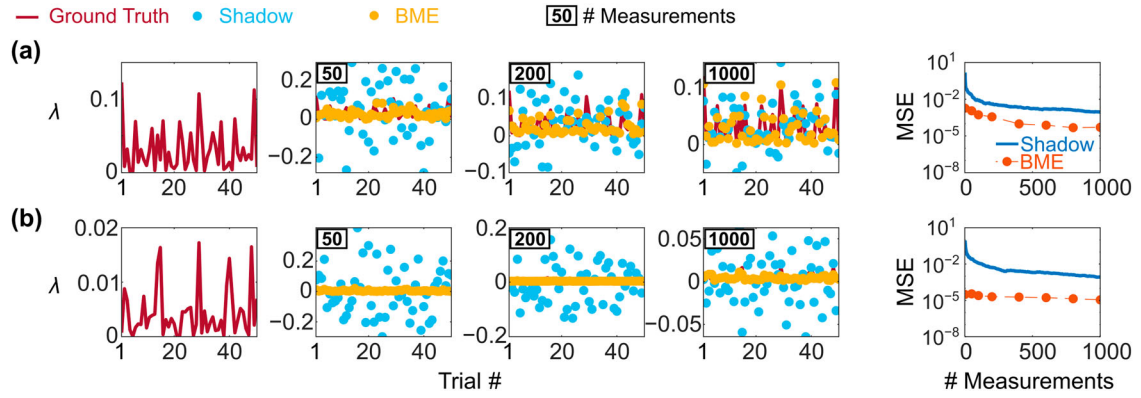
Fig. 2 Estimating rank-1 observable Λ for randomly chosen ground truth states (Picture 2). a $D = 32$ case. **b** $D = 256$ case. The first four columns show λ values for each trial; the last column plots MSE with respect to ground truth over all trials.
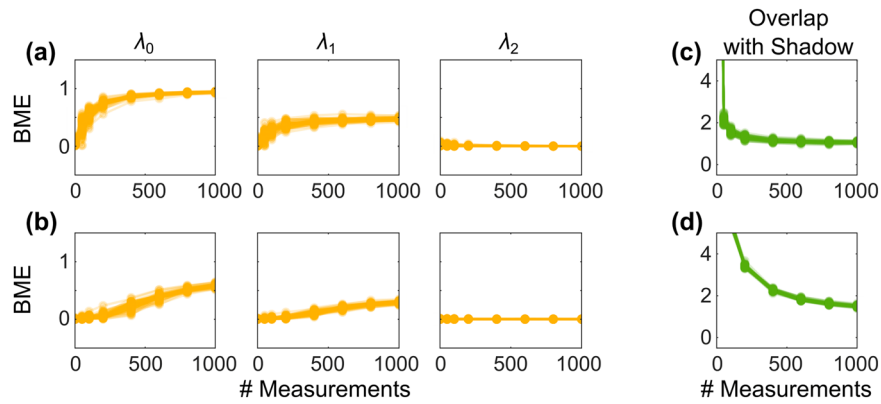

Fig. 3 Bayesian inference results utilizing the pseudo-likelihood in Eq. (9). a $D = 32$ case. **b** $D = 256$ case. The overlap with shadow, $\mathrm{Tr}\,\rho_B\rho_s$ is plotted in **c** for $D = 32$ and **d** for $D = 256$.

measurements, but likely does not care so much about how well the procedure can estimate the ground truth when it is low (e.g., when $\lambda^{(g)} \sim \frac{1}{D}$), since this situation suggests a poorly prepared state anyway. Accordingly, the felt cost is stronger when error is higher for situations with $\lambda^{(g)} \gg \frac{1}{D}$ than when $\lambda^{(g)} \sim \frac{1}{D}$, which is not captured by the standard MSE, as expressed in Eq. (6). And as shown in our tests here, it is precisely these improbable situations wherein shadow excels over BME. Thus, our simulations reveal one surprising reason classical shadows are so powerful: they perform well within those subspaces of the entire Hilbert space that are of interest to a high-fidelity system.

Emulating classical shadows with BME

The dimension independence and rapid convergence of classical shadows for cases of interest indicate the value of a Bayesian version with similar features, both to gain further insight into shadow itself and to improve thereon by ensuring only physically acceptable states. A simple approach for custom Bayesian models, gaining traction in "probably approximately correct" (PAC) learning[27], proposes use of a pseudo-likelihood that rates a prospective state's suitability through a cost function, instead of a full likelihood based on a physical model. In quantum state tomography in particular, quadratic costs of the form $\| \rho - \tilde{\rho} \|_F^2$ have been explored[20,26], where $\tilde{\rho}$ signifies some point estimator and $\| A \|_F = \sqrt{\mathrm{Tr}A^\dagger A}$ the Frobenius norm. Therefore, to obtain a physical state with properties similar to ρ_s, we first suggest the

pseudo-likelihood

$$L_\mathcal{D}(\mathbf{x}) = \exp\left(-\frac{K}{2} \| \rho(\mathbf{x}) - \rho_s \|_F^2 \right). \tag{9}$$

The constant K establishes the relative weight of prior and likelihood. Previously, we suggested $K = M$ for reasonable uncertainty quantification[26]; here we consider $K = MD$ to impart dimension independence. (Incidentally, we have found no significant modifications to the results below when testing with $K \gg MD$.)

Figure 3a, b shows the BME results obtained for $D = 32$ and $D = 256$, respectively, where we again return to Picture 1 with fixed ground truth for all trials. For the tests here, thinning of $T = 2^8$ ($T = 2^{10}$) is used for the $D = 32$ ($D = 256$) MCMC chains. Compared to the shadow results of Fig. 1, the BME predictions still do not reach ground truth values for λ_0 and λ_1 efficiently. This proves intriguing, since $\| \rho - \rho_s \|_F^2$ with $\rho = |\psi\rangle\langle\psi|$ is minimized precisely by states for which $\langle\psi|\rho_s|\psi\rangle$ is large. So if $\lambda_0^{(s)} = \langle g|\rho_s|g\rangle \sim 1$ (cf. Fig. 1), it is odd that predictions using a BME value maximizing $\langle\psi|\rho_s|\psi\rangle$ looks so different for $D = 256$. The origin of this discrepancy, however, lies in ρ_s's nonphysicality.

Plotting the average overlap between shadow and Bayesian samples ($\mathrm{Tr}\,\rho_B\rho_s$) in Fig. 3c, d, we find that ρ_B overlaps with ρ_s more strongly than the ground truth $\rho_g = |g\rangle\langle g|$. Because ρ_s is not positive semidefinite, $\mathrm{Tr}\,\rho_B\rho_s > 1$ for all cases examined. Thus, the BME procedure succeeds in finding states with strong overlap to the shadow, but the closest physical state to ρ_s is not the ground truth, even though $\langle g|\rho_s|g\rangle \sim 1$. Intuitively, this nonphysicality

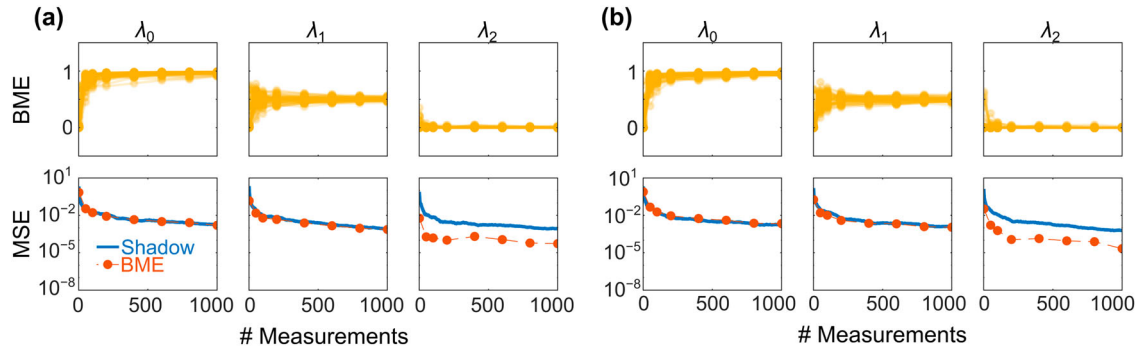**Fig. 4 Bayesian estimation using the pseudo-likelihood of Eq. (10) with $N = 3$. a** Results for $D = 32$. **b** Results for $D = 256$. The MSE values for shadow from Fig. 1 are reproduced for comparison.

helps explain why observables with highly improbable values from the Bayesian view are estimated so much more efficiently with shadow. For a parameterization over physical states and rank-1 observable Λ, only a single state in the Hilbert space attains $\lambda = 1$, and since this represents the maximum value possible for any valid quantum state, it can only be approached from below. On the other hand, a continuum of shadow estimators ρ_s permit $\lambda = 1$, for ρ_s is constrained only by Hermiticity and unit trace—not positive semidefiniteness. Therefore, the estimate $\lambda^{(s)}$ can err on either the high or low side (cf. Fig. 1), pulling the shadow more rapidly to the ground truth in these extreme cases.

This discloses the second central finding of our investigation: the nonphysicality of ρ_s is not a deficiency, but rather critical to obtaining dimension independence. Thus, the key features of the shadow are not necessarily translated onto physical projections like the BME model here or, for that matter, alternative projected-least-squares approaches[28,29]. (As an additional check, we performed the algorithm of ref. [28] to determine the closest physical density matrix to ρ_s, finding very similar results as Fig. 3. This indicates that our projection conclusions are not an artifact of the pure state prior, but hold for general mixed states as well.) While strange from the conventional wisdom of maximum likelihood and BME, nonphysical states are actually beneficial for classical shadows.

Deriving a positive semidefinite model emulating classical shadows remains an intriguing question, however, to eliminate unphysical estimates while retaining the favorable scaling features. With projecting directly onto ρ_s proving unfruitful to this end, we note that, indeed, ρ_s was never intended to serve as an accurate substitute for the true ρ_g; instead, it facilitates estimates of observables[13]. Accordingly, we propose the "observable-oriented pseudo-likelihood"

$$L_{\mathcal{D}}(\mathbf{x}) = \exp\left(-\frac{K}{2}\sum_{n=0}^{N-1}\left|\text{Tr}\,\rho(\mathbf{x})\Lambda_n - \lambda_n^{(s)}\right|^2\right), \quad (10)$$

where we insert the estimates $\lambda_n^{(s)}$ of N observables from ρ_s. This formalism ensures only physical values are returned [through the prior $\pi_0(\mathbf{x})$], and rates the fitness of proposed states through their overlap with respect to shadow's predictions of observables only. For dimension independence, we again set $K = MD$ and perform BME for all simulated datasets and $N = 3$ above, thinning to $T = 2^{10}$ ($T = 2^{13}$) for $D = 32$ ($D = 256$).

The results follow in Fig. 4. Now BME shows very similar behavior to shadow: the MSE with respect to the ground truth matches shadow results from Fig. 1 closely, though BME still outperforms for λ_2. Yet unlike shadow, BME here always gives physically permissible estimates ($\lambda_n^{(B)} \in [0, 1]$). This pseudo-likelihood therefore attains the goal of a BME model commensurate with classical shadows.
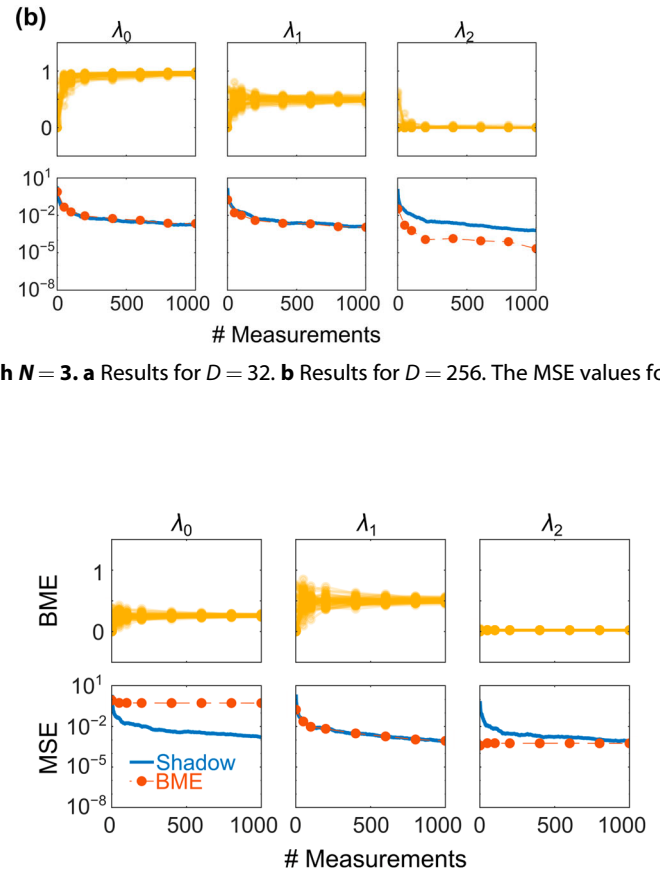

Fig. 5 Bayesian inference results employing the psuedo-likelihood in Eq. (11), for $D = 32$. The shadow MSE values from Fig. 1 are reprinted for clarity.

Yet it is important to emphasize that this approach depends heavily on the quality of the classical shadow. It refines estimates from the shadow with its positive semidefinite requirement, but it does not do markedly better at estimating the ground truth state —at least for arbitrary observables. As an example, we repeat the inference procedure for an observable-oriented pseudo-likelihood based solely on Λ_1, i.e.,

$$L_{\mathcal{D}}(\mathbf{x}) = \exp\left(-\frac{K}{2}\left|\text{Tr}\,\rho(\mathbf{x})\Lambda_1 - \lambda_1^{(s)}\right|^2\right), \quad (11)$$

which has ground truth value $\lambda_2^{(g)} = \frac{1}{2}$. Results for the $D = 32$ case appear in Fig. 5, where we plot the Bayesian estimates for all three observables even though the psuedo-likelihood is based on λ_1 only. The estimate $\lambda_1^{(B)}$ closely matches shadow as designed, and $\lambda_2^{(B)}$ agrees with the ground truth well, due to the fact its value is highly probable for a uniform prior. But $\lambda_0^{(B)} \to \sim \frac{1}{4}$, far from $\lambda_0^{(g)} = 1$.

When using the pseudo-likelihood above, all quantum states with identical overlap to Λ_1 are equally probable, of which the ground truth ρ_g represents just one possibility. The estimate of λ_0 given only λ_1 information reflects the inherent uncertainty within this specification. So to summarize, our observable-oriented pseudo-likelihood builds physicality into shadow, yet it can only (in general) accurately predict the N observables injected into it: to infer quantities beyond these N can prove unreliable.

Highly informative prior

In the above attempts to mimic classical shadows with BME, we focused on pseudo-likelihoods that explicitly leaned on shadow,

either through ρ_s itself [Eq. (9)] or through a set of shadow-estimated observables [Eqs. (10) and (11)]. However, rather than modifying the likelihood (which is based on the physical considerations of Born's rule), a conceptually more satisfying approach would be to modify the prior $\pi_0(\mathbf{x})$ such that any assumptions within classical shadows can be fully incorporated in the BME process without explicitly invoking the unphysical state ρ_s or quantities derived from it.

Unfortunately, while such a shadow-mimetic prior may certainly exist, we have been unable to find one. This objective is complicated by the fact that, even though classical shadows excel over BME with a uniform prior in estimating observables in particular regions of the state space, these regions are not readily defined a priori. These subspaces of high-performance change depending on the particular observable of interest, despite the fact that calculation of ρ_s [Eq. (1)] is independent of the observable to be estimated from it later. Therefore, we suggest one should not view the shadow as biasing a prior distribution; rather, it is working off of an entirely different cost function than MSE [Eq. (6)] —a cost function which assigns variable weights according to the ground truth value of the observable. Research into the nature of this cost optimization would be a fascinating subject for future work.

In the meantime, however, we briefly explore through example how highly informative prior distributions in BME can produce some similar features as classical shadows, albeit only qualitatively. Suppose that one has strong beliefs that the quantum system under test consists of an ideal state $|0\rangle$, which then experiences noisy evolution modeled as a simple depolarizing channel, so that the final quantum state can be expressed as

$$\rho(a) = a|0\rangle\langle 0| + \frac{1-a}{D}I_D, \tag{12}$$

where $a \in [0, 1]$ defines the strength of the noise process. Significantly, this assumption has eliminated the Hilbert space dimension dependence of the number of unknowns: regardless of D, only a single parameter a must be inferred. As a natural choice for simulation, we select the uniform prior $\pi_0(a) = \mathbf{1}_{[0, 1]}(a)$ and consider the full likelihood of Eq. (3), which reduces to

$$L_{\mathcal{D}}(a) = \prod_{m=1}^{M}\left(a|\langle 0|\psi_m\rangle|^2 + \frac{1-a}{D}\right). \tag{13}$$

Utilizing this simplified model, we perform MCMC integration to obtain $\rho_B = \frac{1}{Z}\int da\, L_{\mathcal{D}}(a)\pi_0(a)\rho(a)$ for all previously examined cases under Picture 1; the approach is identical to previous MCMC runs, except (i) thinning was not used to reach convergence in this one-parameter model, and (ii) the proposal distribution for uniform random variables in ref. [30] was applied to maintain the pCN condition. Our findings appear in Fig. 6. Similar to the results

with the observable-oriented pseudo-likelihood in Fig. 4, the BME estimates converge to the ground truth rapidly in the number of measurements, remain in the physically allowable range [0, 1], and show no dependence on dimension D.

These results are obtained without any reference to or reliance on the classical shadow ρ_s. Nonetheless, we emphasize that, despite the extremely high accuracy, these findings build on strong assumptions about the form of the quantum state [Eq. (12)], which are not present in ρ_s. This example highlights, however, that certain features of ρ_s—such as dimension independence—follow automatically from traditional BME as well, under the conditions of a sufficiently informative prior. Of course, whether or not such a simple prior is justified in a particular case depends entirely on one's trust in the quantum system and the goals of the experiment; the salient point is that BME can incorporate these naturally into its formalism.

DISCUSSION
Our numerical investigations here have elucidated two fascinating features of classical shadows:

1. Classical shadows perform extremely well at predicting "unlikely" observables, i.e., those which obtain high values only on a restricted subset of states within the complete Hilbert space.
2. The nonphysicality of classical shadows is critical to their dimension independence and accuracy under few measurements.

These findings do not contradict the optimality of Bayesian methods expressed in Eq. (6): BME with a full likelihood minimizes MSE for any number and collection of measurements, provided the prior distribution accurately reflects the true knowledge involved. The predictive power of ρ_s, then, derives from the fact that the situations in which it is much more accurate that BME are often of particular interest in practice, such as verification of a high-fidelity or highly entangled quantum state. Desiring to extend these features in the Bayesian context, we proposed an observable-oriented pseudo-likelihood that attains shadow's dimension independence and state-specialized accuracy, with the advantage of guaranteed physicality.

Nonetheless, in all these explorations there remains one prominent sense, in which classical shadows unquestionably eclipse BME: computational efficiency. The shadow estimator ρ_s is formed directly from measurements for any dimension D; yet computing ρ_B requires tedious MCMC methods, with the number of parameters increasing linearly (quadratically) with D for a pure (mixed) uniform prior. Here, we considered up to $D = 256$, a far cry from the $D = 2^{120}$ example in ref. [13]. Moving forward, it would
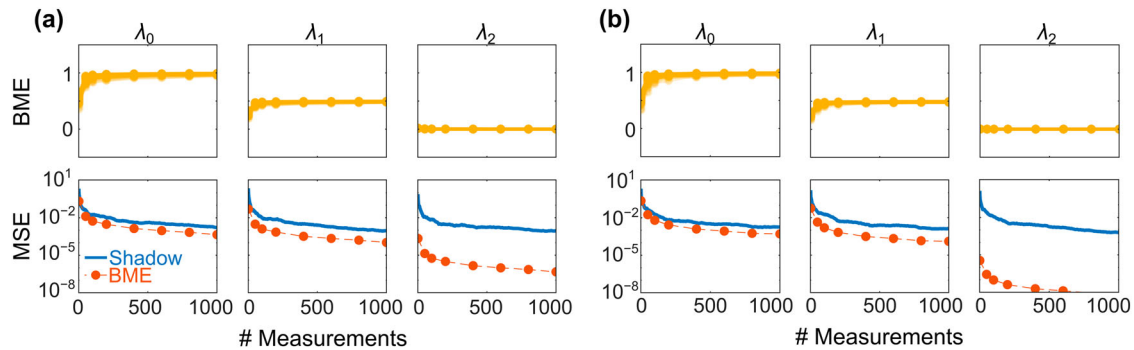

Fig. 6 BME findings utilizing a one-parameter density matrix parameterization [Eq. (12)] and full likelihood [Eq. (13)]. a Results for $D = 32$. **b** Results for $D = 256$. The shadow MSE values from Fig. 1 are plotted again for comparison.

therefore seem profitable to explore simplified Bayesian models that maintain a fixed parameter dimensionality even as the Hilbert space grows exponentially, but that are perhaps not as restrictive as the example in Eq. (12). For example, if one could specify a prior and likelihood on an observable λ only, to the effect of $\pi(\lambda) \propto L_{\mathcal{D}}(\lambda)\pi_0(\lambda)$, the inference procedure would not be limited directly by exponentially large Hilbert spaces. In this way, Bayesian methods could be extendable to the types of quantum systems sought for practically useful quantum computation.

Overall, our analyses have revealed the value of BME as a tool for shedding light on estimation procedures that formally have no connection to the Bayesian paradigm. The numerical simulations here reveal the complementary strengths of classical shadow and Bayesian tomographic approaches in the efficient estimation of quantum properties. And so we expect valuable opportunities for both methods, as quantum information processing resources continue to mature in size and complexity.

METHODS

Data simulation approach

The method of classical shadows introduced in ref. [13] involves application of a Haar-random (or effectively Haar-random) unitary U followed by measurement in the computational basis. We exploit the fact that our target state is pure to substantially reduce the complexity of simulating this procedure. In particular, our simulation method requires the generation of only size-D random vectors rather than $D \times D$ random unitaries.

Without loss of generality, we work in a rotated basis such that the first basis state coincides with the ground truth: $\rho_g = |0\rangle\langle 0|$. Then the probability of observing outcome j depends only on $|\langle j|U|0\rangle|^2 = |U_{j0}|^2 = |(U^{\dagger})_{0j}|^2$. That is, the distribution of outcomes depends only on the first row of $U^{\dagger}$. Now, when U is Haar-random, each individual row and column of $U^{\dagger}$ is a uniformly distributed length-1 vector u. Furthermore, given any component u_j, the remaining components are a uniformly distributed vector of length $\sqrt{1-|u_j|^2}$. A uniformly random vector u, corresponding to the first row of $U^{\dagger}$, may be obtained by generating D complex normal random values and normalizing them to yield a unit length vector. An outcome $n \in \{0, 1, \dots, D-1\}$ is then chosen with probability $|u_n|^2$. This selects the nth column of $U^{\dagger}$. Since this column (whichever it is) is uniformly distributed, its remaining elements are uniformly distributed with length $\sqrt{1-|u_n|^2}$. The explicit procedure is as follows:

1. Posit a measurement unitary $U_m^{\dagger} = [\tilde{\varphi}_0 \cdots \tilde{\varphi}_{D-1}]$, where each $\tilde{\varphi}$ is a column vector corresponding to one of the D possible output states.
2. Generate D complex normal samples $w_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) + i\mathcal{N}(0,1)$ and normalize

$$u_n = \frac{w_n}{\sqrt{\sum_{n'=0}^{D-1}|w_{n'}|^2}}. \tag{14}$$

These define projections of the unitary's basis states on the ground truth: $u_n = \langle 0|\tilde{\varphi}_n\rangle$, or in other words, the elements in the first row of $U_m^{\dagger}$.
3. Select an integer $n \in \{0, 1, \dots, D-1\}$ at random with probability $|u_n|^2$. This implies that the state $\tilde{\varphi}_n$ is detected.
4. Generate $D-1$ complex normal samples $v_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) + i\mathcal{N}(0,1)$ ($j = 1, 2, \dots, D-1$). These set the remaining coefficients of the detected state $\tilde{\varphi}_n$.
5. Finally, take

$$|\psi_m\rangle = u_n|0\rangle + \sqrt{\frac{1-|u_n|^2}{\sum_{j'=1}^{D-1}|v_{j'}|^2}}\sum_{j=1}^{D-1}v_j|j\rangle \tag{15}$$

as the measured state.

Utilizing this method, we performed 50 independent trials with 1000 measurements each, for Hilbert space dimensions $D = 32$ and $D = 256$, giving a total of 100 datasets which are used in all subsequent tests. The

two values of D were selected specifically to clarify how classical shadows and BME differ in their scaling with dimension.

MCMC convergence metrics

For each MCMC chain of length RT, we retained a fixed number of $R = 2^{10}$ sample density matrices $\rho_r = \rho(\mathbf{x}_r)$—or $\rho_r = \rho(a_r)$ for the parameterization in Eq. (12)—and increased the thinning factor T until convergence was confirmed; in practice, we did this by successively doubling T, monitoring the mean of the overlap with the ground truth $\frac{1}{R}\sum_{r=1}^{R}\langle g|\rho_r|g\rangle$, and selecting a T value beyond which no discernible differences in this mean were observed. For each test value of T, we reran the entire chain with a randomly chosen initial condition, which provided a safeguard against the possibility of getting stuck in a local mode. We found the final T value separately for each qualitative regime (likelihood/prior/dimension). This method was simple and allowed us to focus on the quantities of interest (Bayesian means of observables) when gauging convergence. Here, we provide details on the MCMC chains and their convergence properties, in order to show that this approach is consistent with findings from other standard convergence metrics.

All conclusions in this paper depend on estimates of observable values $\Lambda_n = |\phi_n\rangle\langle\phi_n|$, and not the underlying quantum state, so we concentrate on λ_0, λ_1, and λ_2 only [see state definitions in Eq. (7)]. Considering the density matrix MCMC samples $\{\rho_1, \dots, \rho_R\}$, we look at (i) the trace plot $\lambda_n[r]$ ("trace" in MCMC parlance, not matrix trace Tr), (ii) cumulative mean $\mu_n[r]$, and (iii) sample autocorrelation function (ACF) $c_n[l]$, defined as

$$\lambda_n[r] = \langle\phi_n|\rho_r|\phi_n\rangle$$
$$\mu_n[r] = \frac{1}{r}\sum_{k=1}^{r}\lambda_n[k] \tag{16}$$
$$c_n[l] = \frac{1}{C}\sum_{k=1}^{R-l_{\max}}(\lambda_n[k]-\mu_n[R])(\lambda_n[k+l]-\mu_n[R]),$$

where $n \in \{0, 1, 2\}$, $r \in \{1, 2, \dots, R\}$, $l \in \{0, 1, \dots, l_{\max}\}$, and C is a normalization constant such that $c_n[0] = 1$. Of these expressions, the raw trace $\lambda_n[r]$ and autocorrelation $c_n[l]$ probe the independence of the samples in the thinned chain, while the cumulative mean $\mu_n[r]$ is linked to the Bayesian mean estimator, which follows directly from it according to

$$\lambda_n^{(B)} = \mu_n[R]. \tag{17}$$

A total of 3600 separate MCMC chains were required for the BME results in this paper, one chain for each examined combination of likelihood $L_{\mathcal{D}}$, prior distribution π_0, dimension D, number of measurements M, and simulated experimental trial. In lieu of examining all 3600 chains, we exploit the fact that they can be classified into nine qualitative regimes, distinguished by $L_{\mathcal{D}}$, π_0, and D. Figures 7 and 8 plot convergence metrics for one example from each regime, after applying the appropriate thinning value T mentioned in the "Results" section above; for all cases, $l_{\max} = 100$. For each example, we utilize the most extreme case of $M = 1000$ measurements, because this has the most strongly peaked likelihood and therefore requires the longest chains to converge (since the pCN proposal keeps the prior invariant, the acceptance probability varies more strongly as the likelihood becomes more peaked). Thus, a value of T that attains convergence for $M = 1000$ will also realize convergence for $M < 1000$ cases as well.

Overall, the metrics for all nine examples are consistent with converged chains: the trace plots do not reveal long-term trends indicative of correlated points, and the ACFs drop off rapidly with (thinned) lag l. Some features perhaps indicate larger values of T would be preferable, such as the slight modulation in the λ_0 trace of Fig. 8b or the slower drop in ACF in Fig. 8c. Yet such refinements would not be expected to impact the main results, for it is only the final value of the cumulative mean $\mu_n[R]$, which is utilized in comparing to classical shadows. The cumulative means are extremely steady, converging well before $r = R$ for all examples, thus confirming the reliability of the $\lambda_n^{(B)}$ values reported in Figs. 1–6.

Marginal distributions

For additional insight into the joint posterior behavior of $(\lambda_0, \lambda_1, \lambda_2)$—including potential multimodality and correlations—we examine pairwise and single marginal distributions of a subset of the MCMC chains above. We focus on the $D = 256$ case of Fig. 7a and the $D = 32$ case in Fig. 8b, as these show the highest uncertainties. Utilizing kernel density
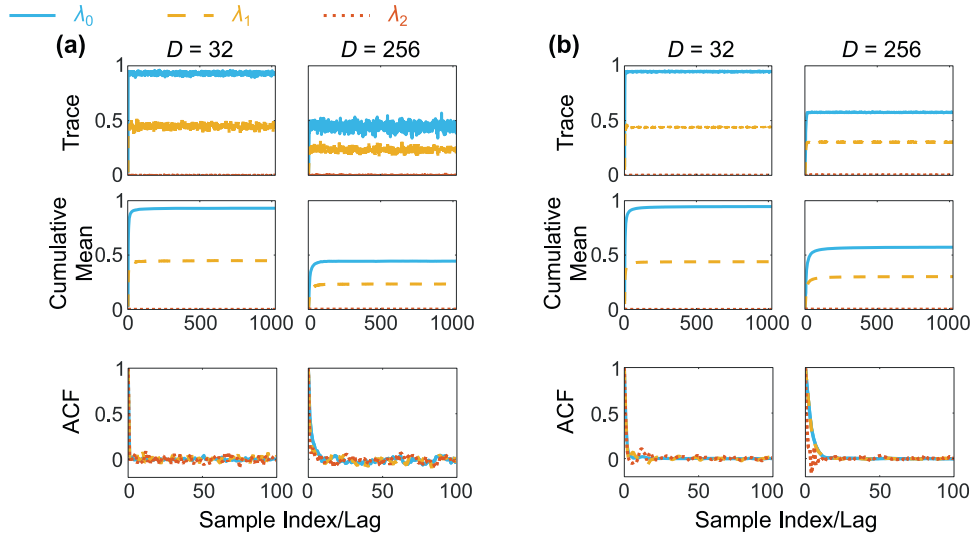
Fig. 7 Example MCMC chains. a Full likelihood of Eq. (3). **b** Pseudo-likelihood of Eq. (9). The three metrics of interest are plotted separately for each observable λ_n. The chains in **a** are examples from Fig. 1 of the main text, and those in **b** from Fig. 3.
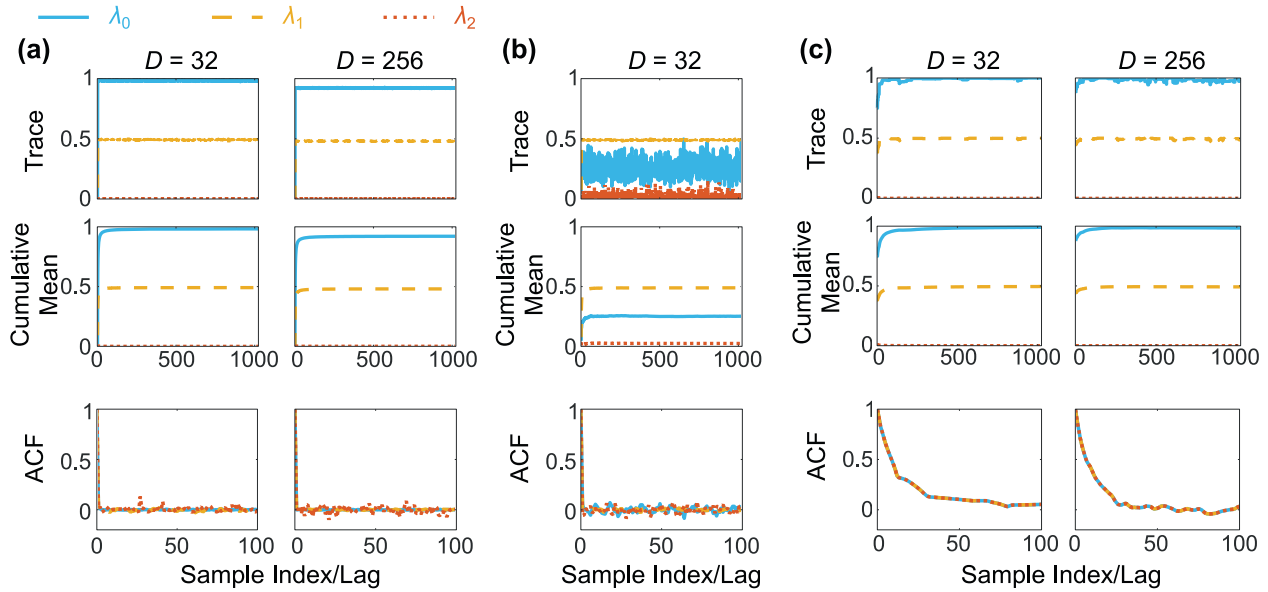

Fig. 8 Example MCMC chains. a Pseudo-likelihood of Eq. (10). **b** Pseudo-likelihood of Eq. (11). **c** Likelihood of Eq. (13). Three metrics are plotted separately for each observable λ_n. The chains in **a** are examples from Fig. 4 of the main text, those in **b** from Fig. 5, and those in **c** from Fig. 6.

estimation in MATLAB[31], and discarding a burn-in of ten samples (to remove the nonstationary initial points evident in the trace plots), we obtain the probability densities in Fig. 9; the two-dimensional contour plots show the pairwise marginals, with the corresponding single marginals appearing to the top and side. All observables follow unimodal distributions. No strong correlations between λ_2 and any other observables are evident from these plots. λ_0 and λ_1 show positive correlations in Fig. 9a, but minimal correlations in Fig. 9b, which can be explained intuitively via the differences in likelihoods. Case (a) corresponds to use of a full likelihood [Eq. (3)] so that the state estimate is gradually moving from $\rho_B = \frac{1}{D}I_D$ (the prior mean, valid for $M \approx 0$) to $\rho_B = |0\rangle\langle 0|$ (the value as $M \to \infty$). Thus, for samples closer to the limit of $\rho_B = |0\rangle\langle 0|$, the values of both $\lambda_0^{(B)}$ and $\lambda_1^{(B)}$ are higher—and vice versa for samples closer to $\rho_B = \frac{1}{D}I_D$. On the other hand, the case in

Fig. 9b corresponds to a pseudo-likelihood based on λ_1 only [Eq. (11)]. Rather than trending toward the ground truth $|0\rangle\langle 0|$, the estimate approaches a continuum of states satisfying $\lambda_1 \sim \frac{1}{2}$, but with no constraint on λ_0, thereby eliminating strong correlations between the two observables.

The fact that the marginals are unimodal, as opposed to multimodal, is not required for any of the theoretical or numerical conclusions in this paper: the results and optimality conditions rely on the Bayesian mean, irrespective of the underlying distribution. However, the unimodality does indicate our pCN technique[26] is well-suited to sampling in this case. It is well known that random walk MCMC methods, of which pCN is an example[25], struggle to reach stationarity under highly multimodal probability densities; in such cases, hybrid Monte Carlo[32] or population-based methods, such as SMC[33], are known to mix much more efficiently.
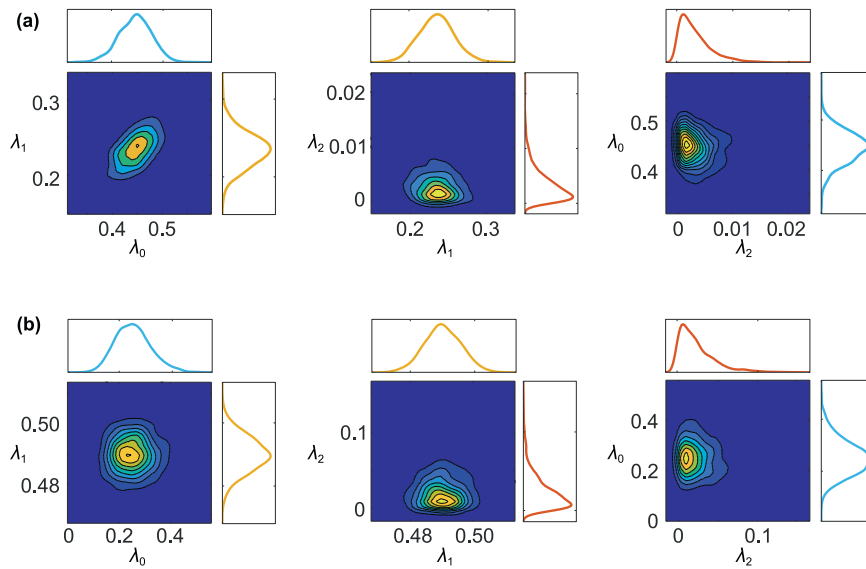
Fig. 9 Pairwise and single marginals for observables λ_0, λ_1, and λ_2. a Points taken from the full likelihood, $D = 256$ example of Fig. 7a. **b** Points taken from pseudo-likelihood, $D = 32$ example of Fig. 8b.

Thus, while pCN would have still applied to multimodal observables, the required values of T—and hence computational time—would have likely increased significantly.

DATA AVAILABILITY

Data used in this study are available from the corresponding author on request.

CODE AVAILABILITY

Analysis code used in this study is available from the corresponding author on request.

REFERENCES

1. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
2. Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**, 150401 (2010).
3. Flammia, S. T., Gross, D., Liu, Y.-K. & Eisert, J. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.* **14**, 095022 (2012).
4. Huszár, F. & Houlsby, N. M. T. Adaptive Bayesian quantum tomography. *Phys. Rev. A* **85**, 052120 (2012).
5. Kravtsov, K. S., Straupe, S. S., Radchenko, I. V., Houlsby, N. M. T., Huszár, F. & Kulik, S. P. Experimental adaptive Bayesian tomography. *Phys. Rev. A* **87**, 062122 (2013).
6. Granade, C., Ferrie, C. & Flammia, S. T. Practical adaptive quantum tomography. *New J. Phys.* **19**, 113017 (2017).
7. Cramer, M. et al. Efficient quantum state tomography. *Nat. Commun.* **1**, 149 (2010).
8. Tóth, G., Wieczorek, W., Gross, D., Krischek, R., Schwemmer, C. & Weinfurter, H. Permutationally invariant quantum tomography. *Phys. Rev. Lett.* **105**, 250403 (2010).
9. Schwemmer, C. et al. Experimental comparison of efficient tomography schemes for a six-qubit state. *Phys. Rev. Lett.* **113**, 040503 (2014).
10. Torlai, G., Mazzola, G., Carrasquilla, J., Troyer, M., Melko, R. & Carleo, G. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).
11. Carrasquilla, J., Torlai, G., Melko, R. G. & Aolita, L. Reconstructing quantum states with generative models. *Nat. Mach. Intell.* **1**, 155–161 (2019).
12. Lohani, S., Kirby, B. T., Brodsky, M., Danaci, O. & Glasser, R. T. Machine learning assisted quantum state estimation. *Mach. Learn. Sci. Technol.* **1**, 035007 (2020).
13. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
14. Aaronson, S. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, 325–338 (ACM, 2018).
15. Blume-Kohout, R. Optimal, reliable estimation of quantum states. *New J. Phys.* **12**, 043034 (2010).
16. Robert, C. P. & Casella, G. *Monte Carlo Statistical Methods* (Springer, 1999).
17. Webb, Z. The Clifford group forms a unitary 3-design. *Quantum Inf. Comput.* **16**, 1379–1400 (2016).
18. Zhu, H. Multiqubit Clifford groups are unitary 3-designs. *Phys. Rev. A* **96**, 062336 (2017).
19. Orenstein, P. Robust mean estimation with the Bayesian median of means. Preprint at https://arxiv.org/abs/1906.01204 (2019).
20. Mai, T. T. & Alquier, P. Pseudo-Bayesian quantum tomography with rank-adaptation. *J. Stat. Plan. Inference* **184**, 62–76 (2017).
21. Seah, Y.-L., Shang, J., Ng, H. K., Nott, D. J. & Englert, B.-G. Monte Carlo sampling from the quantum state space. II. *New J. Phys.* **17**, 043018 (2015).
22. Granade, C., Combes, J. & Cory, D. G. Practical Bayesian tomography. *New J. Phys.* **18**, 033024 (2016).
23. Williams, B. P. & Lougovski, P. Quantum state estimation when qubits are lost: a no-data-left-behind approach. *New J. Phys.* **19**, 043003 (2017).
24. Lu, H.-H. et al. A controlled-NOT gate for frequency-bin qubits. *npj Quantum Inf.* **5**, 24 (2019).
25. Cotter, S. L., Roberts, G. O., Stuart, A. M. & White, D. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.* **28**, 424–446 (2013).
26. Lukens, J. M., Law, K. J. H., Jasra, A. & Lougovski, P. A practical and efficient approach for Bayesian quantum state estimation. *New J. Phys.* **22**, 063038 (2020).
27. Guedj, B. A primer on PAC-Bayesian learning. Preprint at https://arxiv.org/abs/1901.05353 (2019).
28. Smolin, J. A., Gambetta, J. M. & Smith, G. Efficient method for computing the maximum-likelihood quantum state from measurements with additive Gaussian noise. *Phys. Rev. Lett.* **108**, 070502 (2012).
29. Guță, M., Kahn, J., Kueng, R. & Tropp, J. A. Fast state tomography with optimal error bounds. *J. Phys. A Math. Theor.* **53**, 204001 (2020).
30. Vollmer, S. J. Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors. *SIAM/ASA J. Uncertain. Quantif.* **3**, 535–561 (2015).
31. MathWorks. "ksdensity" www.mathworks.com/help/stats/ksdensity.html (2021).
32. Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216–222 (1987).
33. Del Moral, P., Doucet, A. & Jasra, A. Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B* **68**, 411–436 (2006).

AUTHOR CONTRIBUTIONS

J.M.L. performed the numerical simulations and led writing of the paper. K.J.H.L. provided guidance on Bayesian methods. R.S.B. developed the overall research question and supervised the project. All authors furnished ideas, analyzed the results, and contributed to the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to J.M.L.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.