



Enhancing quantum state reconstruction with structured classical shadows



Zhen Qin¹, Joseph M. Lukens^{2,3,4}, Brian T. Kirby^{5,6} & Zhihui Zhu¹ ✉

While classical shadows can efficiently predict key quantum state properties, their suitability for certified quantum state tomography remains uncertain. In this paper, we address this challenge by introducing a projected classical shadow (PCS) that extends the standard classical shadow by incorporating a projection step onto the target subspace. For a general quantum state consisting of n qubits, our method requires a minimum of $O(4^n)$ total state copies to achieve a bounded recovery error in the Frobenius norm between the reconstructed and true density matrices, reducing to $O(2^n r)$ for states of rank $r < 2^n$ —meeting information-theoretic optimal bounds in both cases. For matrix product operator states, we demonstrate that the PCS can recover the ground-truth state with $O(n^2)$ total state copies, improving upon the previously established Haar-random bound of $O(n^3)$. Numerical simulations validate our scaling results and demonstrate the practical accuracy of the proposed PCS method.

Quantum state tomography (QST) is widely used for estimating quantum states^{1–5}. To reconstruct the density matrix with high accuracy, measurements should be performed on a large number of identical copies; specifically, for single-copy (i.e., non-collective) measurements, a minimum of $O(4^n)$ total copies is required to estimate the density matrix of an n -qubit system with a bounded recovery error, as defined by the Frobenius norm between the reconstructed and true density matrices⁶. Various methods have been proposed to achieve efficient and accurate QST. Classical computational approaches include linear inversion⁷, maximum likelihood estimation^{4,5,8}, Bayesian inference^{9–11}, region estimation^{12,13}, classical machine learning¹⁴, and least squares estimators^{15–17}. In contrast, quantum machine learning methods encompass algorithms such as variational quantum circuits^{18,19}, quantum principal component analysis²⁰, and quantum variational algorithms combined with classical statistics²¹.

A significant reduction in the number of required state copies can be achieved by assuming two common low-dimensional structures: low-rankness and matrix product operators (MPOs). (i) Low-rank density matrices frequently emerge in quantum systems with pure or nearly pure states that exhibit low entropy^{6,16,22–24}, and low-rank assumptions are employed in various state estimation procedures, with a range of associated measurement processes, including 4-designs²², Pauli strings^{23,25}, Clifford gates¹⁶, and Haar-random projective measurements²⁴. When the density matrix has rank r , the required number of total state copies can be reduced to $O(2^n r)$ ^{6,16}, yet this remains exponential in n , posing challenges for current quantum computers exceeding 100 qubits. (ii) MPOs, on the other hand, offer a more scalable alternative for certain quantum systems, including

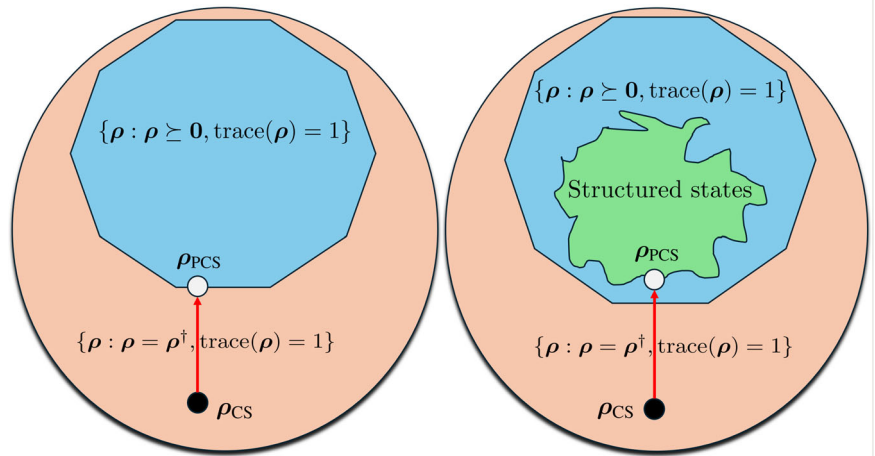
one-dimensional spatial systems²⁶, Hamiltonians with decaying long-range interactions²⁷, and states generated by noisy quantum devices²⁸. When employing Haar-random projective measurements²⁹ or specific classes of informationally complete positive operator-valued measures (IC-POVMs)³⁰, the required number of total state copies can be reduced to polynomial scaling—either $O(n^3)$ or $O(n)$, respectively—while ensuring bounded recovery error for MPO states.

While algorithms with low-rank assumptions or low-dimensional structures can enable significantly improved scaling, they still face considerable computational complexity, which in existing approaches can be attributed to four potential operations: (i) the calculation of the inverse; (ii) repeated inner product operations between matrices that grow exponentially with n ; (iii) multiple projection steps onto the target subspace; and (iv) additional matrix multiplications introduced by iterative algorithms to enforce low-rankness or MPO representations. Recently, an efficient and experimentally feasible approach, known as classical shadow (CS) estimation, was introduced by ref. 31 to infer limited sets of state properties like fidelity, entanglement measures, and correlations. By exploiting efficient computational and storage capabilities on classical hardware, all necessary processing to predict these properties can be carried out via classical computations. This has sparked a series of studies leveraging the CS method^{32–36}. Meanwhile, CS has also been utilized for full quantum state reconstruction^{37,38}, and integrated with projection techniques to recover physical quantum states in^{39,40}. Yet to the best of our knowledge, no existing theoretical analysis of the sampling complexity for the CS-based method addresses the reconstruction of other structured quantum states, above and

¹Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA. ²Elmore Family School of Electrical and Computer Engineering and Purdue Quantum Science and Engineering Institute, Purdue University, West Lafayette, IN, USA. ³Quantum Information Science Section, Oak Ridge National Laboratory, Oak Ridge, TN, USA. ⁴Research Technology Office and Quantum Collaborative, Arizona State University, Tempe, AZ, USA. ⁵DEVCOM Army Research Laboratory, Adelphi, MD, USA. ⁶Tulane University, New Orleans, LA, USA. ✉ e-mail: zhu.3440@osu.edu

Fig. 1 | Illustration of proposed PCS method.

Given an initial CS estimate ρ_{CS} lying in the space of Hermitian and unit-trace matrices (not necessarily PSD), we compute the closest state ρ_{PCS} in the physical space of interest—either the space of all possible states (left) or a subspace possessing a desired structure (right).



beyond simply enforcing physicality. Thus, a theoretical understanding of whether the CS-based method can be effectively extended to the *full state* (i.e., QST) with provable performance guarantees remains absent.

In this paper, we derive performance guarantees for QST using a method we term projected classical shadow (PCS), which projects CS estimators onto target subspaces of the Hilbert space, as illustrated in Fig. 1. Given that the original CS density matrix is Hermitian but not in general positive semidefinite (PSD), our method involves projecting its eigenvalues onto the simplex⁴¹. We demonstrate that this approach requires $O(4^n)$ total state copies to sufficiently achieve a bounded recovery error in the Frobenius norm. For low-rank states, we further leverage (truncated) low-rank eigenvalue decomposition and show that the required number of total state copies can be reduced to $O(2^n r)$ for the same accuracy. Finally, for MPO states, we employ a quasi-optimal MPO projection—tensor-train singular value decomposition (TT-SVD)⁴² with a simplex projection—to form the PCS step, demonstrating that with $O(n^2)$ total state copies, the method reliably recovers the ground-truth state. While suboptimal relative to the degrees of freedom for MPO states, this approach improves upon the theoretical $O(n^3)$ scaling in ref. 29. PCS also offers a framework for incorporating prior knowledge about the target state form into the CS approach.

Notation

We use bold capital letters (e.g., $\mathbf{X}$) to denote matrices, bold lowercase letters (e.g., $\mathbf{x}$) to denote column vectors, and italic letters (e.g., x) to denote scalar quantities. Matrix elements are denoted in parentheses. For example, $X(i_1, i_2)$ denotes the element in position (i_1, i_2) of the matrix $\mathbf{X}$. The superscripts $(\cdot)^T$ and $(\cdot)^\dagger$ denote the transpose and Hermitian transpose, respectively. For two matrices $\mathbf{A}, \mathbf{B}$ of the same size, $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}^\dagger \mathbf{B})$ denotes the inner product. $\|\mathbf{X}\|$, $\|\mathbf{X}\|_1$, and $\|\mathbf{X}\|_F$ respectively represent the spectral, trace, and Frobenius norm of $\mathbf{X}$. For two positive quantities $a, b \in \mathbb{R}^+$, the inequality $b \lesssim a$ or $b = O(a)$ implies $b \leq ca$ for some universal constant c ; likewise, $b \gtrsim a$ or $b = \Omega(a)$ represents $b \geq ca$ for some universal constant c .

Results

Classical shadows

Quantum information science harnesses quantum states for information processing⁴³. The state of an n -qubit system can be described by the density operator $\rho \in \mathbb{C}^{2^n \times 2^n}$, which is PSD ($\rho \geq 0$) and has unit-trace ($\text{trace}(\rho) = 1$). In order to estimate this state, measurements can be performed on a collection of copies.

Projective measurements. Within the most general quantum measurement framework of positive operator valued measures (POVMs) (Specifically, a POVM is characterized as a set of PSD matrices: $\{\mathbf{A}_1, \dots, \mathbf{A}_K\} \in \mathbb{C}^{2^n \times 2^n}$, s. t. $\sum_{k=1}^K \mathbf{A}_k = \mathbf{I}_{2^n}$. Each POVM element $\mathbf{A}_k$

corresponds to a potential outcome of a quantum measurement with the special case of projective measurements corresponding to the case where all $\mathbf{A}_k$ are pairwise orthogonal projection operators, meaning they satisfy $\mathbf{A}_k^2 = \mathbf{A}_k$ and $\mathbf{A}_k \mathbf{A}_j = 0$ for $k \neq j$.), the special case of projective measurements is often employed, where the measurement outcomes are associated with an orthonormal eigenbasis of the system. To implement such a measurement defined by an arbitrary orthonormal basis $\{\phi_k : \phi_k^\dagger \phi_l = \delta_{kl}\}$, we can introduce a unitary matrix $\mathbf{U} = [\phi_1 \dots \phi_{2^n}] \in \mathbb{C}^{2^n \times 2^n}$ and apply $\mathbf{U}^\dagger$ to the state ρ before conducting a projective measurement in the computational basis $\{e_k\}$, where $\mathbf{U}e_k = \phi_k$. The probability of observing the k -th outcome is given by:

$$p_k = \langle \phi_k | \rho | \phi_k \rangle = e_k^\dagger (\mathbf{U}^\dagger \rho \mathbf{U}) e_k. \quad (1)$$

However, a single projective measurement, even if repeated infinitely many times, provides only partial information on ρ , so multiple projective measurements must be conducted in various bases. In the subsequent discussion, we denote the number of distinct measurement bases by M , and the measurement operators for the m -th projective measurement by $\{\phi_{m,1}^\dagger \phi_{m,1}, \dots, \phi_{m,2^n}^\dagger \phi_{m,2^n}\}$.

Classical shadow (CS). Consider the original CS proposal with single-shot Haar-random projective measurements. Given an unknown n -qubit ground truth ρ^* , we repeatedly execute the measurement procedure above Eq. (1) in which $\mathbf{U}$ is chosen randomly from the Haar distribution and each measurement is performed on only one copy (i.e., a new $\mathbf{U}$ is selected for each copy measured). The specific result e_{j_m} yields a snapshot, or “shadow,” of the underlying quantum state, which for Haar-distributed unitaries can be expressed as³¹:

$$\begin{aligned} \rho_m &= (2^n + 1) \mathbf{U}_m e_{j_m} e_{j_m}^\dagger \mathbf{U}_m^\dagger - \mathbf{I}_{2^n} \\ &= (2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}. \end{aligned} \quad (2)$$

By construction, this snapshot equals the ground truth in expectation (over both unitaries and measurement outcomes): $\mathbb{E}[\rho_m] = \rho^*$. Executing this process M times produces an array of M independent classical snapshots for the total CS estimator:

$$\begin{aligned} \rho_{\text{CS}} &= \frac{1}{M} \sum_{m=1}^M \rho_m \\ &= \frac{1}{M} \sum_{m=1}^M [(2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}]. \end{aligned} \quad (3)$$

CS for Tomography? Although CS estimators can efficiently predict observables of ρ^* , to our knowledge, there exist no theoretical results concerning the recovery error of the *full state*. Following the detailed derivation in the section “Methods”, we find the expectation of the mean squared error:

$$\mathbb{E} \|\rho_{\text{CS}} - \rho^*\|_F^2 = \frac{4^n + 2^n - 1 - \|\rho^*\|_F^2}{M}. \quad (4)$$

Given that $\|\rho^*\|_F^2 \leq [\text{trace}(\rho^*)]^2 = 1$, it follows that Eq. (4) can be simplified to

$$\mathbb{E} \|\rho_{\text{CS}} - \rho^*\|_F^2 \approx \frac{4^n}{M} \quad (5)$$

for large n . Eq. (5) demonstrates that stable recovery of the full state can be achieved only when M scales proportionally to 4^n , aligning with the optimal M required in QST for general states⁶.

A comparison between CS and traditional QST returns several key observations of relevance to this study:

1. CS yields an unbiased estimate ($\mathbb{E}[\rho_{\text{CS}}] = \rho^*$), whereas the solution from QST is often biased⁴⁴, due to the fact that most QST methods involve physical constraints, such as positivity and unit trace.
2. While the CS estimator is typically unphysical (not PSD), leading QST methods like maximum likelihood estimation (MLE)⁵, projected least squares⁴⁵, and Bayesian inference⁹ enforce physicality by construction.
3. CS boasts significantly lower computational complexity compared to standard QST methods, such as MLE using the fixed-point (FP) algorithm^{46,47}, MLE using gradient descent (GD)⁴⁸, least squares (LS) using GD²⁹, and the one-step LS method¹⁷. The number of iterations required for convergence in iterative methods [e.g., MLE (FP), MLE (GD), and LS (GD)] significantly increases computational complexity. Moreover, the requirement of a suitable initialization further imposes a strong and often nontrivial condition for successful recovery. Although the one-step LS method¹⁷ avoids iterations, it does not incorporate any constraint and still involves matrix inversion and multiple matrix multiplications, resulting in substantial computational cost. In contrast, the CS method offers the lowest computational complexity, as it is also a one-step approach whose primary cost arises from computing the outer product of two vectors.
4. For $M \ll 2^n$, CS outperforms QST in predicting certain linear observables, not in predicting the entire state^{17,49}.
5. Including prior information about state structure allows for a reduction in scaling in QST (see Tables 1 and 2). Apart from specialized CS methods tailored for states generated by shallow circuits and Hamiltonian dynamics^{50,51}, which aim to improve the accuracy of predicting quantum state properties, there currently exists no known approach that similarly reduces the sample complexity of CS for full quantum state reconstruction. In other words, CS requires $O(4^n)$ measurements for estimating the full state, as demonstrated in Eq. (5).

In the next section we investigate methods for incorporating prior information about state structure into CS to reduce the scaling shown in Eq. (5).

Projected Classical Shadow (PCS) for QST

In this section, we will study the application of CS for the task of describing the full quantum state and show that, with a simple projection step, CS estimators are also effective for QST and achieve (nearly) information-theoretically optimal bounds for broad classes of states. Let $\mathbb{X}$ denote the class of states of interest, and assume that the underlying ground truth $\rho^* \in \mathbb{X}$. For instance, $\mathbb{X}$ could contain all physical states (PSD and unit-trace) or be restricted to a specific structure with compact representations, such as low-rank or MPO states. Because of the availability of previous sample complexity results based on the Frobenius norm, we choose to define

Table 1 | Comparing the total number of state copies in PCS using single-shot global Haar unitaries to that in optimal QST

Methods	$\rho^* \in \mathbb{X}_{\text{simplex}}$	$\rho^* \in \mathbb{X}_r$
PCS	$O(8^n/\xi^2)$	$O(2^{n^2}/\xi^2)$
Optimal QST ⁶	$\Omega(8^n/\xi^2)$	$\Omega(2^{n^2}r^2/(\xi^2 \log(1/\xi)))$

Here, n denotes the number of qubits, r represents the rank of the target state, and ξ signifies the desired precision in trace distance, i.e., $\|\hat{\rho} - \rho^*\|_1 \leq \xi$ for the estimator $\hat{\rho}$.

Table 2 | Total number of copies in MPO-PCS compared to constrained LS using Haar measures, and spherical 3-designs

Method	$\rho \in \mathbb{X}_D$
Approximate PCS (Haar)	$O(n^2 D^2 \log n / \zeta^2)$
Optimal PCS (Haar)	$O(n D^2 \log n / \zeta^2)$
Constrained LS (Haar) ²⁹	$O(n^3 D^2 \log n / \zeta^2)$
Constrained LS (spherical 3-designs) ³⁰	$O(n D^2 \log n / \zeta^2)$

Here, n denotes the number of qubits, and ζ signifies the desired precision in Frobenius distance, i.e., $\|\hat{\rho} - \rho^*\|_F \leq \zeta$ for the estimator $\hat{\rho}$.

ρ_{PCS} as the projection of ρ_{CS} on the set $\mathbb{X}$ that minimizes Frobenius error, i.e.,

$$\rho_{\text{PCS}} = \mathcal{P}_{\mathbb{X}}(\rho_{\text{CS}}) := \arg \min_{\rho \in \mathbb{X}} \|\rho - \rho_{\text{CS}}\|_F. \quad (6)$$

To provide a unified and general analysis of Eq. (6), we enlist tools from ϵ -net and covering number^{29,52} to capture the complexity of the classes of states within the set $\mathbb{X}$. First, consider the set $\mathcal{N} = \left\{ \frac{\rho}{\|\rho\|_F} : \rho \in \mathbb{X} \right\}$ scaled to unit Frobenius norm. For $\epsilon > 0$, the set $\mathcal{N}_\epsilon \subset \mathcal{N}$ is said to be an ϵ -net (or an ϵ -cover) over $\mathcal{N}$ if for all $\frac{\rho}{\|\rho\|_F} \in \mathcal{N}$, there exists $\frac{\rho'}{\|\rho'\|_F} \in \mathcal{N}_\epsilon$ such that $\left\| \frac{\rho}{\|\rho\|_F} - \frac{\rho'}{\|\rho'\|_F} \right\|_F \leq \epsilon$. The size of an ϵ -net with the smallest cardinality is called the covering number of $\mathbb{X}$, denoted by $N_\epsilon(\mathbb{X})$. Intuitively speaking, a covering number is the minimum number of balls of a specified radius ϵ to cover a given set entirely. Coverings are useful for managing the complexity of a large set: instead of directly analyzing the behavior of an uncountable number of points in $\mathcal{N}$, we can analyze the finite number of points in $\mathcal{N}_\epsilon$. The behavior of all points in $\mathcal{N}$ is similar to that of the points in $\mathcal{N}_\epsilon$, as each point in $\mathcal{N}$ is close to some point in the covering.

Instead of the covering number $N_\epsilon(\mathbb{X})$, our analysis will rely on the covering number of the set $\overline{\mathbb{X}}$ formed by the differences between the elements in $\mathbb{X}$:

$$\overline{\mathbb{X}} = \{\rho_1 - \rho_2 : \rho_1, \rho_2 \in \mathbb{X}, \rho_1 \neq \rho_2\}. \quad (7)$$

In many cases, the covering number $N_\epsilon(\overline{\mathbb{X}})$ can be upper-bounded by $N_\epsilon^2(\mathbb{X})$. Here we use $\overline{\mathbb{X}}$ for convenience in the following.

The covering number when $\mathbb{X}$ comprises all physical quantum states can be computed as $\log N_\epsilon(\overline{\mathbb{X}}) = O(4^n \log \frac{2}{\epsilon})$. By comparison, for quantum states with rank at most r , this reduces to $\log N_\epsilon(\overline{\mathbb{X}}) = O(2^n r \log \frac{2}{\epsilon})$; when the density matrices are represented by MPOs with bond dimension D , the covering number can be further reduced to $\log N_\epsilon(\overline{\mathbb{X}}) = O(4n D^2 \log \frac{4n+\epsilon}{\epsilon})$, as discussed in the next part.

Theorem 1. For a given $\rho^* \in \mathbb{X}$, let ρ_{PCS} be the projected CS in Eq. (6). Then with probability at least $1 - e^{-\Omega(\log N_{1/2}(\mathbb{X}))}$,

$$\|\rho_{\text{PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{\log N_{1/2}(\mathbb{X})}{M}}\right). \quad (8)$$

The proof is given in the section “Methods”. Here, the set $\mathbb{X} \subset \{\rho \in \mathbb{C}^{2^n \times 2^n} : \rho = \rho^\dagger, \text{trace}(\rho) = 1\}$ is any subspace of Hermitian, trace-one matrices (tan space in Fig. 1). The set $\mathbb{X}$ will be specialized to PSD matrices only (blue space in Fig. 1) in Corollary 1 and low-dimensional structures (green space in Fig. 1) in Theorems 3 and 4. Theorem 1 guarantees a stable recovery of the ground-truth ρ^* with ξ -closeness in the Frobenius norm, provided $M = O(\log N_{1/2}(\mathbb{X})/\xi^2)$ number of Haar-random projective measurements, which scales linearly with the logarithm of the covering number. For structured sets $\mathbb{X}$ that are nonconvex, such as MPO states, computing the optimal projection $\mathcal{P}_{\mathbb{X}}$ might be difficult or even NP-hard. For these cases, we can use numerical methods to compute an approximate projection $\tilde{\mathcal{P}}_{\mathbb{X}}$ that we assume is α -approximately optimal ($\alpha \geq 1$), satisfying

$$\tilde{\mathcal{P}}_{\mathbb{X}}(\rho) \in \mathbb{X}, \quad \|\tilde{\mathcal{P}}_{\mathbb{X}}(\rho) - \rho\|_F \leq \sqrt{\alpha} \|\mathcal{P}_{\mathbb{X}}(\rho) - \rho\|_F \quad (9)$$

for any ρ . As will be explained in the next sections, there exist efficient methods to find approximation projections for low-rank and MPO states. Denote by $\tilde{\rho}_{\text{PCS}} = \tilde{\mathcal{P}}_{\mathbb{X}}(\rho_{\text{CS}})$ the PCS estimator obtained with this approximate projection. The following extends the results in Theorem 1 to $\tilde{\rho}_{\text{PCS}}$.

Theorem 2. For a given $\rho^* \in \mathbb{X}$, let $\tilde{\rho}_{\text{PCS}}$ be the approximate PCS estimator in Eq. (9). Then with probability at least $1 - e^{-\Omega(\log N_{1/2}(\mathbb{X}))}$,

$$\|\tilde{\rho}_{\text{PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{\alpha \log N_{1/2}(\mathbb{X})}{M}}\right). \quad (10)$$

General physical states. We first specialize $\mathbb{X}$ to all physical quantum states (We chose the label “simplex” for this set since the eigenvalues $\{\lambda_k\}$ of all physical states define a standard simplex, i.e., $\lambda_k \geq 0$ and $\sum_k \lambda_k = 1$):

$$\mathbb{X}_{\text{simplex}} = \{\rho \in \mathbb{C}^{2^n \times 2^n} : \rho \succcurlyeq 0, \text{trace}(\rho) = 1\}. \quad (11)$$

For $\mathbb{X}_{\text{simplex}}$, the PCS projection in Eq. (6) can be implemented by performing an eigenvalue decomposition, followed by projecting the eigenvalues onto the simplex $\{x \in \mathbb{R}^{2^n} : x_i \geq 0, \sum_{i=1}^{2^n} x_i = 1\}$ using the algorithm proposed in Refs. 41,45, while keeping the eigenvectors unchanged. This approach has also been employed in Refs. 39,40 to ensure the physical structure of the reconstructed state. The computational complexity of the projection step is $O(a \log a)$, where a denotes the number of nonzero eigenvalues.

Since the corresponding set $\mathbb{X}$ has covering number $\log N_{\epsilon}(\mathbb{X}) = O(4^n \log \frac{2}{\epsilon})$, we can plug this information into Theorem 1 to obtain recovery guarantee for $\mathcal{P}_{\text{simplex}}(\rho_{\text{CS}})$.

Corollary 1. For a given physical state $\rho^* \in \mathbb{C}^{2^n \times 2^n}$, we perform M projective measurements to obtain the CS estimate ρ_{CS} . Then with probability at least $1 - e^{-\Omega(4^n)}$, the projected classical shadow $\mathcal{P}_{\text{simplex}}(\rho_{\text{CS}})$ satisfies

$$\|\mathcal{P}_{\text{simplex}}(\rho_{\text{CS}}) - \rho^*\|_F \leq O\left(\sqrt{\frac{4^n}{M}}\right). \quad (12)$$

Low-rank states. We next explore the structure of pure or nearly pure quantum states characterized by low entropy and represented as low-rank density matrices. Assuming ρ^* has rank $r \leq 2^n$, we can refine our attention to the set $\mathbb{X}_r = \{\rho \in \mathbb{C}^{2^n \times 2^n} : \rho \succcurlyeq 0, \text{trace}(\rho) = 1, \text{rank}(\rho) = r\}$. Denote $\mathcal{P}_{\mathbb{X}_r}(\cdot)$ as the optimal projection satisfying Eq. (6). It follows from Theorem 1 and the covering number of the corresponding set $\log N_{\epsilon}(\mathbb{X}_r) = O(2^n r \log \frac{2}{\epsilon})$ that $\|\mathcal{P}_{\mathbb{X}_r}(\rho_{\text{CS}}) - \rho^*\|_F \leq O(\sqrt{2^n r/M})$.

However, since we are unaware of an algorithm to perform the ideal projection $\mathcal{P}_{\mathbb{X}_r}(\cdot)$, we instead consider a two-step alternative to obtain the low-rank projected classical shadow (LR-PCS):

$$\rho_{\text{LR-PCS}} = \mathcal{P}_{\text{simplex}}(\mathcal{P}_{\text{rank-}r}(\rho_{\text{CS}})), \quad (13)$$

where $\mathcal{P}_{\text{rank-}r}(\cdot)$ denotes the rank- r projection obtained by setting all eigenvalues beyond the r -th largest eigenvalue to zero. We can show that $\rho_{\text{LR-PCS}}$ shares a similar guarantee as $\mathcal{P}_{\mathbb{X}_r}(\rho_{\text{CS}})$.

Theorem 3. Given M Haar-random projective measurements on physical state $\rho^* \in \mathbb{X}_r$, with probability $1 - e^{-\Omega(2^n r)}$ $\rho_{\text{LR-PCS}}$, defined in Eq. (13), satisfies

$$\|\rho_{\text{LR-PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{2^n r}{M}}\right). \quad (14)$$

The detailed proof appears in the section “Methods”. This theoretical recovery error is optimal, given that the degrees of freedom for the ground truth ρ^* are $O(2^n r)$. This highlights that LR-PCS can achieve the optimal solution in QST using independent measurements, without requiring multiple iterations of optimization algorithms.

To compare LR-PCS with prior results, we convert the result of Theorem 3 to trace norm leveraging the inequality between the Frobenius and the trace norms⁶, namely $\|\rho_{\text{LR-PCS}} - \rho^*\|_1 \leq \sqrt{2r} \|\rho_{\text{LR-PCS}} - \rho^*\|_F \leq O(\sqrt{2^n r^2/M})$, which matches the optimal guarantee (up to small log terms) with independent measurements according to ref. 6. We have summarized the comparison in Table 1. We note that the sufficient condition for PCS in the general setting matches the necessary condition established in ref. 6. Similarly, the sufficient condition for PCS in the low-rank setting is also close to the corresponding necessary condition in ref. 6, up to logarithmic factors.

MPO states. While the computational and storage requirements for low-rank density matrices are significantly smaller compared to general ones, they still grow exponentially in the number of qubits n . Moreover, the assumption of high purity on which the low-rank approximation is based becomes increasingly tenuous in practice for existing processors in the noisy intermediate-scale quantum (NISQ) era. For this reason, reducing parameter count through alternative assumptions is worth pursuing. Examples such as ground states of many quantum systems with short-range interactions and states generated by such systems within a finite duration²⁶ often possess entanglement localized to subsystems of the entire quantum computer. Consequently, they can be compactly represented using MPOs, whose degrees of freedom scale only polynomially in n . To assist in the development of an MPO-PCS method, we will first establish their connection to tensor train (TT) decompositions⁴², a technique widely utilized in signal processing and machine learning.

For a n -qubit density matrix $\rho^* \in \mathbb{C}^{2^n \times 2^n}$, we employ a single index array $i_1 \dots i_n$ ($j_1 \dots j_n$) to denote the row (column) indices, where $i_1, \dots, i_n \in \{1, 2\}$ (Specifically, $i_1 \dots i_n$ represents the $(i_1 + \sum_{\ell=2}^n 2^{\ell-1}(i_\ell - 1))$ -th row.). We designate ρ^* as an MPO if we can represent its $(i_1 \dots i_n, j_1 \dots j_n)$ -th element using the following matrix product⁵³:

$$\rho^*(i_1 \dots i_n, j_1 \dots j_n) = X_1^{i_1 j_1} X_2^{i_2 j_2} \dots X_n^{i_n j_n}, \quad (15)$$

where $X_\ell^{i_\ell j_\ell} \in \mathbb{C}^{D \times D}$ for $\ell \in \{2, \dots, n-1\}$, $X_1^{i_1 j_1} \in \mathbb{C}^{1 \times D}$, $X_n^{i_n j_n} \in \mathbb{C}^{D \times 1}$, and D is the bond dimension, and thus we can introduce the set of physical

MPO states with bond dimension D as

$$\mathbb{X}_D = \left\{ \rho \in \mathbb{C}^{2^n \times 2^n} : \rho \succcurlyeq 0, \text{trace}(\rho) = 1, \text{bond dimension}(\rho) = D \right\}. \quad (16)$$

Here, the corresponding difference set $\overline{\mathbb{X}}$ has a covering number $\log N_\epsilon(\overline{\mathbb{X}}) = O(4nD^2 \log \frac{4n+e}{\epsilon})$, which is proportional to the degrees of freedom $O(4nD^2)$ in the MPO states. Given the optimal projection $\mathcal{P}_{\mathbb{X}_D}(\cdot)$, it follows from Theorem 1 that $\|\mathcal{P}_{\mathbb{X}_D}(\rho_{\text{CS}}) - \rho^*\|_F \leq O(\sqrt{nD^2 \log n/M})$.

However, we have been unable to implement the optimal $\mathcal{P}_{\mathbb{X}_D}(\cdot)$ due to the difficulty in satisfying both the MPO and simplex conditions simultaneously. Therefore, we introduce a quasi-optimal projection based on a sequential singular value decomposition (SVD) algorithm, commonly referred to as tensor train SVD (TT-SVD)⁴². Based on tensor-matrix equivalence, we can design a two-step MPO PCS method:

$$\rho_{\text{MPO-PCS}} = \mathcal{P}_{\text{simplex}}(\text{SVD}_D^{\text{tt}}(\rho_{\text{CS}})), \quad (17)$$

where $\text{SVD}_D^{\text{tt}}(\cdot)$ denotes the TT-SVD operation. It is worth noting that the bond dimension of $\rho_{\text{MPO-PCS}}$ may differ slightly from D due to the simplex projection, but the recovery error still depends on D . We analyze the recovery error of Eq. (17) as follows:

Theorem 4. Consider an MPO state $\rho^* \in \mathbb{X}_D$, measured M times with Haar-random projections. With $\rho_{\text{MPO-PCS}}$ defined as in Eq. (17), with probability $1 - e^{-\Omega(nD^2 \log n)}$ we have

$$\|\rho_{\text{MPO-PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{n^2 D^2 \log n}{M}}\right). \quad (18)$$

The proof can be found in the section “Methods”. Note that due to the quasi-optimality of the TT-SVD, the upper bound of Eq. (18) is not optimal when considering the degrees of freedom $O(4nD^2)$ in ρ^* . To our knowledge,

Table 3 | Average runtime per trial (reported as mean $\pm$ standard deviation, in seconds) of numerical experiments in Fig. 4, for the case $M = 1000$

Method	Thermal ($T = 0.2$)	Thermal ($T = 2$)	GHZ
CS	$(5.30 \pm 0.24) \times 10^{-3}$	$(5.80 \pm 0.51) \times 10^{-3}$	$(5.20 \pm 0.11) \times 10^{-3}$
LR-PCS	$(1.50 \pm 0.17) \times 10^{-2}$	$(1.42 \pm 0.25) \times 10^{-2}$	$(1.44 \pm 0.11) \times 10^{-2}$
MPO-PCS	$(3.16 \pm 0.52) \times 10^{-2}$	$(3.37 \pm 0.85) \times 10^{-2}$	$(2.97 \pm 0.28) \times 10^{-2}$
LR-MLE	49.1 ± 4.4	57.4 ± 2.8	49.7 ± 3.3
MPO-MLE	117.2 ± 4.3	113.3 ± 5.7	112.4 ± 7.7

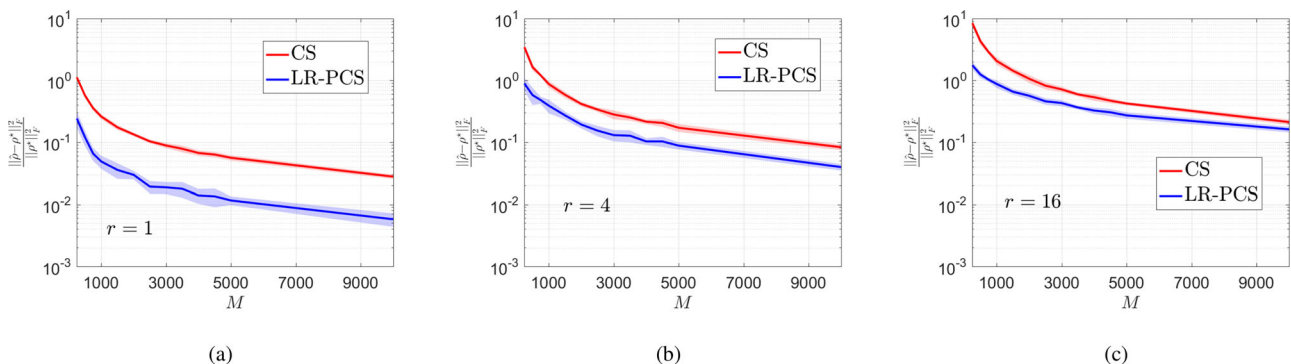


Fig. 2 | Estimating four-qubit low-rank states by CS and LR-PCS methods. Mean squared error as a function of state copies M , averaged over trials on ten randomly chosen ground truth states of rank $r = 1$ (a), $r = 4$ (b), and $r = 16$ (c). The horizontal axes span $M = 250$ to $M = 10000$. Uncertainty is defined as the sample standard deviation.

there exists no method that can guarantee both MPO and PSD constraints simultaneously. Should such an optimal MPO projection be found, however, we could potentially remove one factor of n in the numerator of Eq. (18), thus ensuring exact MPO rank. In Table 2, we summarize the total number of state copies required for MPO-PCS compared to existing QST methods. It is important to highlight that all QST results represent sufficient, rather than necessary, conditions. Compared to the constrained LS method using Haar measurements in ref. 29, MPO-PCS exhibits better sample complexity. While the result³⁰ based on constrained least squares with spherical 3-design POVMs—which are informationally complete—could potentially outperform PCS in terms of sample complexity (but PCS achieves the same complexity given an optimal projection), attaining the bound in ref. 30 requires solving a highly nonconvex optimization problem to global optimality. Table 3. However, there is currently a lack of practical algorithms capable of achieving this bound. The gradient-based iterative algorithm proposed in ref. 30 provides only a suboptimal guarantee and is initialization-dependent—which may limit its practical applicability. Additionally, spherical 3-designs are not known to have efficient implementations using current local quantum circuits, whereas Haar measurements are generally more feasible in experimental settings.

Simulation results

In this section, we conduct numerical QST experiments with Haar-random projective measurements to compare CS, LR-PCS, and MPO-PCS methods. For each configuration, we conduct 10 Monte Carlo tomographic experiments in which each Haar measurement and result are sampled at random; then we take the average over all 10 trials to report the results. For the random state cases (Figs. 2, 3), each trial corresponds to a different randomly chosen ground truth, whereas for the tailored state cases (Figs. 4, 5), each trial in a given average is performed on the same ground truth. Furthermore, since the magnitude of $\|\rho^*\|_F$ differs across quantum states with different ranks or bond dimensions, we apply the normalized Frobenius norm to enable fair comparisons and provide a consistent metric for reconstruction accuracy.

In the first set of tests, we compare CS and LR-PCS for a specific rank r as a function of measurements M . We generate random ground-truth density matrices $\rho^* = F^* F^{*\dagger} \in \mathbb{C}^{16 \times 16}$ ($n = 4$ qubits), where $F^* = \frac{A^* + iB^*}{\|A^* + iB^*\|_F} \in \mathbb{C}^{16 \times r}$, and the entries of A^* and B^* are independent and identically distributed (i.i.d.) samples drawn from the standard normal distribution. Notably, when $r = 16$, LR-PCS reduces to projection onto the set of general physical states defined in Eq. (11). The results in Fig. 2 for rank $r \in \{1, 4, 16\}$ reveal two key observations: (i) as the rank r decreases and the number of measurements M increases, the recovery error across all methods consistently reduces, with the squared error quantitatively scaling as expected ($4^n/M$ for CS and $2^n r/M$ for LR-PCS); (ii) for any r and M , LR-PCS outperforms standard CS (even at full rank), as it preserves physicality under any rank constraints; and (iii) in Fig. 2c, since $r = 16$ (i.e., ρ^* is full rank), LR-PCS provides only the additional PSD constraint compared to CS. As M

Fig. 3 | Estimating seven-qubit MPO states by CS and LR-PCS methods. Mean squared error as a function of state copies M , where each point is the average over trials on ten randomly chosen ground truth states for bond dimension $D = 1$ (a) and $D = 4$ (b). Uncertainty is defined as the sample standard deviation.

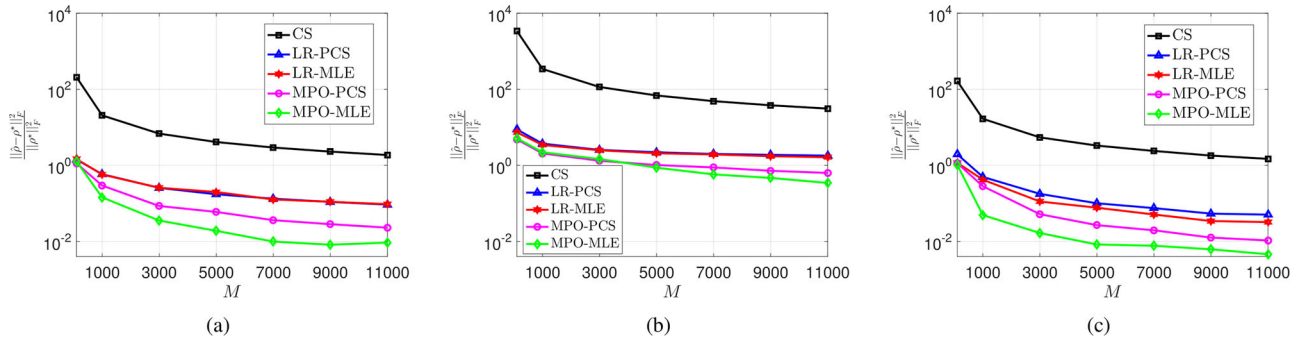
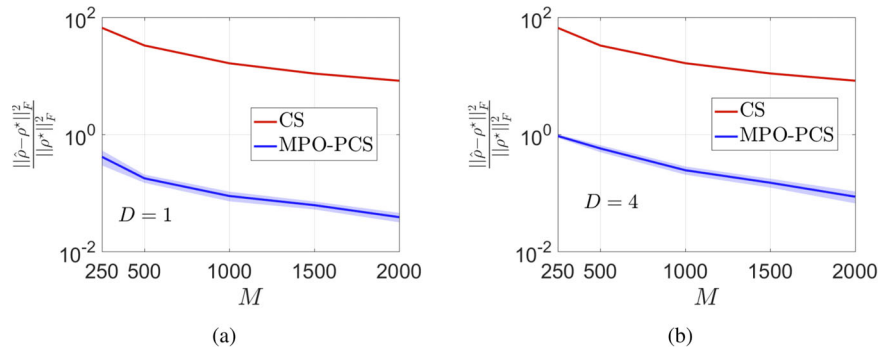


Fig. 4 | Estimating seven-qubit thermal and GHZ states. Mean squared error as a function of the number of state copies M for (a) thermal state ($T = 0.2$), (b) thermal state ($T = 2$), and (c) GHZ state. Comparison between different methods for (a)

thermal state ($T = 0.2$), (b) thermal state ($T = 2$), and (c) GHZ state. All figures have $M = 100$ as the starting point.

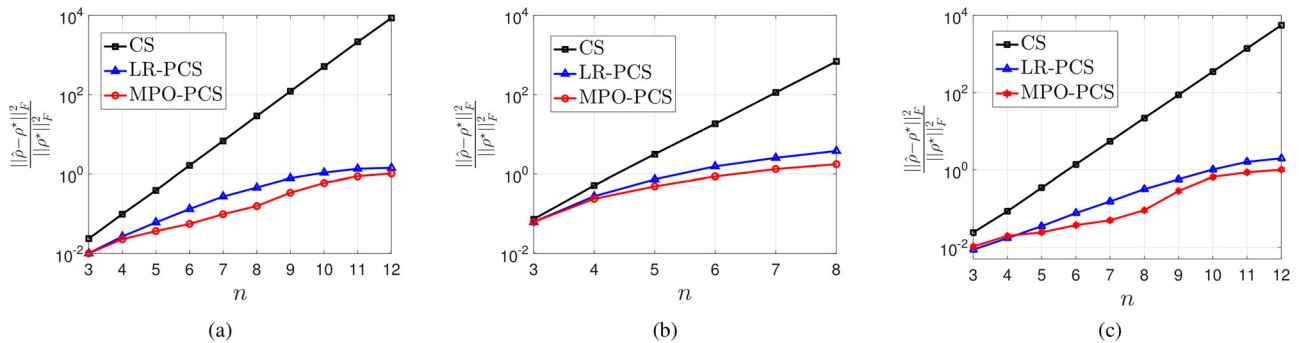


Fig. 5 | Estimating thermal and GHZ states with varied number of qubits. Mean squared error as a function of the total qubit number with $M = 3000$ for (a) thermal state ($T = 0.2$), (b) thermal state ($T = 2$), and (c) GHZ state.

increases, the performance gap between LR-PCS and CS narrows, as the PSD constraint alone has limited impact on reducing the recovery error.

In the second set of trials, we test CS and MPO-PCS across varying numbers of measurements M and bond dimension D . We consider $n = 7$ -qubit matrix product states (MPSs, pure state special cases of MPOs) of the form $\rho^* = \mathbf{u}^* \mathbf{u}^{*\dagger} \in \mathbb{C}^{128 \times 128}$, where $\mathbf{u}^* \in \mathbb{C}^{128 \times 1}$ satisfies $\|\mathbf{u}^*\|_2 = 1$ and its $(i_1 \dots i_7)$ -element can be represented in the matrix product form: $\mathbf{u}^*(i_1 \dots i_7) = \mathbf{U}_1^{*i_1} \dots \mathbf{U}_7^{*i_7}$. Here, each matrix $\mathbf{U}_\ell^{*i_\ell}$ has size $d \times d$, except for $\mathbf{U}_1^{*i_1}$ and $\mathbf{U}_7^{*i_7}$ of dimensions $1 \times d$ and $d \times 1$, respectively.

To generate each MPS $\mathbf{u}^*$, we draw a length-128 complex vector with i.i.d. standard normal elements, apply TT-SVD⁴² to truncate it to an MPS, and then normalize the result to unit length. As a result, entry $\rho^*(i_1 \dots i_7, j_1 \dots j_7)$ can be expressed as $\rho^*(i_1 \dots i_7, j_1 \dots j_7) = (\mathbf{U}_1^{*i_1} \otimes \mathbf{U}_1^{*j_1}) \dots (\mathbf{U}_7^{*i_7} \otimes \mathbf{U}_7^{*j_7}) = \mathbf{X}_1^{*i_1 j_1} \dots \mathbf{X}_7^{*i_7 j_7}$, where $\otimes$ denotes the Kronecker product. Thus, $\rho^* = \mathbf{u}^* \mathbf{u}^{*\dagger}$ is also an MPO with bond dimension $D = d^2$ (equal for all qubits). As shown in Fig. 3, MPO-PCS attains significantly

lower error than CS, as it leverages knowledge about the underlying MPO structure. And the recovery error of MPO-PCS increases with higher MPO bond dimension (in line with Table 2), whereas that of CS remains the same regardless of D .

In the third set of trials, we simulate measurements on 7-qubit density matrices: (i) thermal state (The thermal state is generated from the 1D quantum Ising model $H = \sum_{j=1}^{n-1} \sigma_j^z \sigma_{j+1}^z + \sum_{j=1}^n \sigma_j^x$ with $\sigma_j^a = \mathbf{I}_{2^{j-1}} \otimes \sigma^a \otimes \mathbf{I}_{2^{n-j}} \in \mathbb{R}^{2^n \times 2^n}$, $a = x, z$ and $\sigma^x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\sigma^z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$). The thermal state is then defined as $\rho^* = \frac{e^{-H/T}}{\text{trace}(e^{-H/T})}$ with temperature $T = 0.2$ (a relatively low temperature close to the ground state); (ii) thermal state with temperature $T = 2$ (corresponding to a relatively high temperature); and (iii) Greenberger–Horne–Zeilinger (GHZ) state (The GHZ state is constructed as $\rho^* = \mathbf{g} \mathbf{g}^\dagger$ where $\mathbf{g} = \left[\frac{1}{\sqrt{2}} \ 0 \ \dots \ 0 \ \frac{1}{\sqrt{2}} \right]^\top \in \mathbb{R}^{2^n \times 1}$). It is worth

noting that the low-temperature thermal state (i) and the GHZ state (iii) simultaneously exhibit low-rank and MPO structures^{16,48,54}, making them well-suited for demonstrating the advantages of exploiting structured subspaces. We impose a rank constraint $r \in \{4, 24, 1\}$ for the estimator on each state, respectively. For the $T = 0.2$ thermal state, the ground-truth density matrix has rank of approximately 4, while for the high-temperature case ($T = 2$), it is full-rank; for LR-PCS $r = 24$ is selected, somewhat arbitrarily, which is sufficient to encompass 80% of the sum of the eigenvalues of the ground-truth density matrix. In addition, we apply TT-SVD on the CS estimator to adaptively select the bond dimensions using the error tolerance 10^{-14} . To facilitate a comprehensive comparison between the PCS-based methods and MLE, we introduce the low-rank MLE (LR-MLE) and matrix product operator MLE (MPO-MLE) algorithms, as detailed in the section “Methods”. For LR-MLE and MPO-MLE, we adopt random initialization (We generate a density matrix $\rho_0 = F_0 F_0^\dagger \in \mathbb{C}^{2^n \times 2^n}$, where $F_0 = \frac{A_0 + iB_0}{\|A_0 + iB_0\|_F} \in \mathbb{C}^{2^n \times r}$, with the entries of A_0 and B_0 independently drawn from the standard normal distribution.) and spectral initialization³⁰, respectively. The step sizes are set to 0.5, 0.05, and 0.1 for LR-MLE, and 0.3, 0.05, and 0.1 for MPO-MLE across the three set of quantum states mentioned above; the number of iterations is fixed at 500 to guarantee convergence for all cases. Given that the variance error in reconstructing the thermal state and GHZ state is less than 10% of the mean error, we exclude this error from the figure to preserve its clarity. Fig. 4 shows that the proposed LR-PCS and MPO-PCS methods outperform standard CS, as quantified by the Frobenius norm. Furthermore, MPO-PCS demonstrates superior performance compared to LR-PCS, which can be attributed to the lower degrees of freedom in the MPO structure relative to the low-rank structure [cf. Eqs. (14), (18)]. In addition, LR-PCS achieves performance comparable to that of LR-MLE, as it attains the information-theoretically optimal error bound. In contrast, the performance of MPO-PCS is slightly inferior to that of MPO-MLE, primarily due to its suboptimal recovery error bound, which contains a factor of n^2 rather than n . Nevertheless, according to Table 3, we note that CS-based methods are significantly more computationally efficient than MLE-based methods (more than $100 \times$ faster in the cases considered), as the latter require iterative optimization procedures.

In the final test, we examine how the recovery error scales with qubit number n , using parameter settings of $r = 4$ for $T = 0.2$, $r = 4(n - 1)$ for $T = 2$, $r = 1$ for GHZ state and an error tolerance of 10^{-14} for determining bond dimension D . As highlighted in Fig. 5, both LR-PCS and MPO-PCS effectively attenuate the growth in recovery error as the system size n increases, in contrast to the standard CS method. This improvement is attributed to the utilization of the low-dimensional structure in these methods. Additionally, the recovery error of MPO-PCS scales polynomially with n , as indicated in Eq. (18), rather than exponentially as in Eq. (14) of LR-PCS; hence, MPO-PCS outperforms LR-PCS in terms of recovery error.

Discussion

This paper has introduced the projected classical shadow (PCS) method to address the computational challenges of quantum state tomography (QST) in large Hilbert spaces by leveraging the classical shadow (CS) framework combined with a physical projection step. The method provides guaranteed performance under Haar-random measurements. Theoretical results show that the PCS method achieves high accuracy in reconstructing general and low-rank quantum states while minimizing the number of state copies, meeting information-theoretically optimal bounds. Moreover, the PCS method reduces the number of state copies required for matrix product operator (MPO) states compared to existing results using Haar random measurements. Numerical validation further demonstrates the practicality and computational efficiency of PCS for large-scale quantum state reconstruction.

More broadly, our formalism points to a promising new general direction for CS methods. Although originally introduced for the estimation of state *properties* rather than the state *per se*³¹, CS nevertheless relies on an estimator ρ_{CS} of the full density matrix. As our results reveal, this generally

unphysical estimator can be projected onto a physical space of interest—whether the entire Hilbert space or some subset thereof (Fig. 1)—with performance guarantees that attain information-theoretic bounds (for the case of arbitrary and low-rank states) or improve upon previous scaling results (for MPO states). Therefore in merging the conceptual simplicity of CS with the scaling improvements possible in structured quantum systems, our results suggest a compelling role for PCS in traditional quantum state estimation, with exciting opportunities for future exploration in even more types of subspaces tailored to specific physical conditions or prior knowledge, such as projected entangled pair operator (PEPO)⁵⁵ and multiscale entanglement renormalization ansatz (MERA)⁵⁶ constructions.

Another promising direction is to analyze the PCS method under local measurements. Although global measurements—characterized by joint operations across all qubits—are theoretically advantageous, their implementation using practical quantum circuits poses substantial challenges. In contrast, local measurements—whether taken from the Haar measure¹⁶, the Pauli set²³, or local informationally complete POVMs⁴⁸—are significantly more compatible with current quantum architectures and can be implemented with greater experimental efficiency. While the projection-based framework employed in this work could, in principle, be directly adapted to local measurement scenarios, the theoretical machinery developed herein does not extend to such cases, as the concentration inequality in Eq. (27) [see section “Methods”] cannot be established under local measurements. Addressing this gap necessitates the development of new analytical tools tailored to the locality constraints, which we leave as an important direction for future investigation.

Methods

This section provides detailed proofs and a comprehensive description of the MLE-based methods (LR-MLE and MPO-MLE) introduced in the last section.

Proof of Equation 4

Proof. We expand $\mathbb{E}\|\rho_{CS} - \rho^*\|_F^2$ as follows:

$$\begin{aligned} \mathbb{E}\|\rho_{CS} - \rho^*\|_F^2 &= \mathbb{E}\left\|\frac{1}{M} \sum_{m=1}^M \rho_m - \rho^*\right\|_F^2 \\ &= \mathbb{E}\left\langle \frac{1}{M} \sum_{m=1}^M (\rho_m - \rho^*), \frac{1}{M} \sum_{m=1}^M (\rho_m - \rho^*) \right\rangle \\ &= \frac{1}{M^2} \mathbb{E} \sum_{m=1}^M \|\rho_m - \rho^*\|_F^2 \\ &= \frac{1}{M} \mathbb{E} \|\rho_1 - \rho^*\|_F^2 \\ &= \frac{1}{M} (\|\rho^*\|_F^2 - 2\mathbb{E}\langle \rho_1, \rho^* \rangle + \mathbb{E}\langle \rho_1, \rho_1 \rangle) \\ &= \frac{1}{M} \left[-\|\rho^*\|_F^2 + (2^n + 1)^2 \mathbb{E}\langle \phi_{1,j_1} \phi_{1,j_1}^\dagger, \phi_{1,j_1} \phi_{1,j_1}^\dagger \rangle \right. \\ &\quad \left. - 2(2^n + 1) \mathbb{E}\langle \phi_{1,j_1} \phi_{1,j_1}^\dagger, \mathbf{I}_{2^n} \rangle + 2^n \right] \\ &= \frac{4^n + 2^n - 1 - \|\rho^*\|_F^2}{M}, \end{aligned} \quad (19)$$

where the third line follows from $\mathbb{E}[\rho_m] = \rho^*$, the fourth from the equivalence under expectation of all measurements m , and the last from the normalization $\langle \phi_{1,j_1} \phi_{1,j_1}^\dagger, \phi_{1,j_1} \phi_{1,j_1}^\dagger \rangle = \langle \phi_{1,j_1} \phi_{1,j_1}^\dagger, \mathbf{I}_{2^n} \rangle = 1$.

Proof of Theorem 1

Proof. We define a restricted Frobenius norm as

$$\begin{aligned} \|\rho_{PCS} - \rho^*\|_{F, \widehat{\mathbb{X}}} &= \|\rho_{PCS} - \rho^*\|_F \\ &= \max_{\rho \in \widehat{\mathbb{X}}} \langle \rho_{PCS} - \rho^*, \rho \rangle, \end{aligned} \quad (20)$$

where $\widehat{\mathbb{X}} = \{\rho \in \mathbb{C}^{2^n \times 2^n} : \text{trace}(\rho) = 0, \rho = \rho^\dagger, \|\rho\|_F \leq 1\}$. By the definition of the restricted Frobenius norm in Eq. (20), we can further analyze

$$\begin{aligned} \|\rho_{\text{PCS}} - \rho^*\|_F &= \|\rho_{\text{PCS}} - \rho^*\|_{F, \widehat{\mathbb{X}}} \leq \|\rho_{\text{CS}} - \rho^*\|_{F, \widehat{\mathbb{X}}} \\ &= \max_{\rho \in \widehat{\mathbb{X}}} \left\langle \frac{1}{M} \sum_{m=1}^M \left[(2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} \right] - \rho^*, \rho \right\rangle, \end{aligned} \quad (21)$$

where the inequality follows from the assumption that the physical projection $\mathcal{P}_{\mathbb{X}}(\cdot)$ is optimal and therefore satisfies nonexpansiveness. Next, we bound $\frac{1}{M} \sum_{m=1}^M [(2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}] - \rho^*$ using the covering argument. According to the assumption, we initially construct an ϵ -net $\{\rho^{(1)}, \dots, \rho^{(N_\epsilon(\widehat{\mathbb{X}}))}\} \in \widehat{\mathbb{X}} \subset \widehat{\mathbb{X}}$, where the size of $\widehat{\mathbb{X}}$ is denoted by $N_\epsilon(\widehat{\mathbb{X}})$ such that

$$\sup_{\rho: \|\rho\|_F \leq 1} \min_{\rho \in N_\epsilon(\widehat{\mathbb{X}})} \|\rho - \rho^{(p)}\|_F \leq \epsilon. \quad (22)$$

In addition, we denote $\mathbf{B}_m = \frac{1}{M} ((2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*)$ and derive

$$\begin{aligned} \max_{\rho \in \widehat{\mathbb{X}}} \left\langle \sum_{m=1}^M \mathbf{B}_m, \rho \right\rangle &= \max_{\rho \in \widehat{\mathbb{X}}} \left\langle \sum_{m=1}^M \mathbf{B}_m, \rho - \rho^{(p)} + \rho^{(p)} \right\rangle \\ &\leq \max_{\rho^{(p)} \in \widehat{\mathbb{X}}} \left\langle \sum_{m=1}^M \mathbf{B}_m, \rho^{(p)} \right\rangle + \epsilon \max_{\rho \in \widehat{\mathbb{X}}} \left\langle \sum_{m=1}^M \mathbf{B}_m, \rho \right\rangle. \end{aligned}$$

By setting $\epsilon = 0.5$ and moving the second term on the right-hand side to the left, we get

$$\max_{\rho \in \widehat{\mathbb{X}}} \left\langle \sum_{m=1}^M \mathbf{B}_m, \rho \right\rangle \leq \max_{\rho^{(p)} \in \widehat{\mathbb{X}}} 2 \left\langle \sum_{m=1}^M \mathbf{B}_m, \rho^{(p)} \right\rangle. \quad (23)$$

Then we need to build the concentration inequality for the right-hand side of Eq. (23). First, we define

$$\sum_{m=1}^M s_m = \sum_{m=1}^M \left\langle (2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \rho^{(p)} \right\rangle, \quad (24)$$

and due to $\mathbb{E}[(2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*] = \mathbf{0}$, we have $\mathbb{E}[s_m] = 0$. Moreover, we rewrite s_m as

$$\begin{aligned} s_m &= \left\langle (2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \rho^{(p)} \right\rangle \\ &= (2^n + 1) \left\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger - \frac{\rho^*}{2^n + 1}, \rho^{(p)} \right\rangle \\ &= (2^n + 1) \left\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, \rho^{(p)} \right\rangle - \frac{\langle \rho^*, \rho^{(p)} \rangle}{2^n + 1} \langle \mathbf{I}_{2^n}, \rho^{(p)} \rangle \\ &= (2^n + 1) \left\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, \mathbf{D} \right\rangle, \end{aligned} \quad (25)$$

where the second line follows from $\text{trace}(\rho^{(p)}) = \langle \mathbf{I}_{2^n}, \rho^{(p)} \rangle = 0$. We can further compute

$$\begin{aligned} \mathbb{E}[|s_m|^a] &= \mathbb{E} \left[(2^n + 1)^a \left| \left\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, \mathbf{D} \right\rangle \right|^a \right] \\ &\leq (2^n + 1)^a \mathbb{E} \left[(\text{trace}(\phi_{m,j_m} \phi_{m,j_m}^\dagger |\mathbf{D}|))^a \right] \\ &= \frac{(2^n + 1)^a}{C_{2^n+a-1}} \text{trace}(|\mathbf{D}|^{\otimes a} P_{\text{Sym}}) \\ &\leq \frac{(2^n + 1)^a}{C_{2^n+a-1}} \|\mathbf{D}\|_F^{\otimes a} \|P_{\text{Sym}}\| \\ &\leq 6 \times 2^{a-2} a!, \end{aligned} \quad (26)$$

where $|\mathbf{D}| = \sqrt{\mathbf{D}^2} = \mathbf{U} \sqrt{\Sigma} \mathbf{V}^\dagger$ denotes the absolute value of the matrix $\mathbf{D}$ with its compact SVD $\mathbf{D}^2 = \mathbf{U} \Sigma \mathbf{V}^\dagger$ and $\mathbf{A}^{\otimes a} = \underbrace{\mathbf{A} \otimes \dots \otimes \mathbf{A}}_a$ holds for any

matrix $\mathbf{A}$. Given that the unitary Haar measure conforms to any unitary p -design, as exemplified in [ref. 57, Example 51], we can deduce the third line, with P_{Sym} representing an orthogonal projector onto the symmetric subspace. The second inequality follows from [ref. 58, Lemma 7] and $\|\mathbf{D}\|_F^{\otimes a} = \|\mathbf{D}\|^{\otimes a}_F$ due to the positive semidefiniteness of $|\mathbf{D}|^{\otimes a}$ and the orthogonal projection. In the last line, we utilize $\|\mathbf{D}\|_F \leq \|\rho^{(p)}\|_F + \|\frac{\rho^*}{2^n+1} \mathbf{I}_{2^n}\|_F \leq 1 + \frac{2^n}{2^n+1} \|\rho^{(p)}\|_F \leq 2$, $\|P_{\text{Sym}}\| \leq 1$ and $\frac{(2^n+1)^a}{C_{2^n+a-1}} \leq \frac{3}{2} a!$.

Based on Lemma 1 with $\mathbb{E}[s_m] = 0$ and $\mathbb{E}[|s_m|^a] \leq 6 \times 2^{a-2} a!$, for any $t \in [0, 1]$, we have the probability

$$\begin{aligned} \mathbb{P} \left(\left| \frac{1}{M} \sum_{m=1}^M \left\langle (2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \rho^{(p)} \right\rangle \right| \geq t \right) \\ \leq 2e^{-\frac{Mt^2}{28}}. \end{aligned} \quad (27)$$

Combining Eqs. (23), (27), there exists an ϵ -net $\widetilde{\mathbb{X}}$ of $\widehat{\mathbb{X}}$ such that

$$\begin{aligned} \mathbb{P} \left(\max_{\rho \in \widetilde{\mathbb{X}}} \left\langle \frac{1}{M} \sum_{m=1}^M \left[(2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} \right] - \rho^*, \rho \right\rangle \geq t \right) \\ \leq \mathbb{P} \left(\max_{\rho^{(p)} \in \widetilde{\mathbb{X}}} \frac{1}{M} \sum_{m=1}^M \left\langle (2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \rho^{(p)} \right\rangle \geq \frac{t}{2} \right) \\ \leq \mathbb{P} \left(\max_{\rho^{(p)} \in \widetilde{\mathbb{X}}} \frac{1}{M} \sum_{m=1}^M \left| \left\langle (2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \rho^{(p)} \right\rangle \right| \geq \frac{t}{2} \right) \\ \leq 2N_\epsilon(\widetilde{\mathbb{X}}) e^{-\frac{Mt^2}{112}} \\ \leq e^{-\frac{Mt^2}{112} + \log 2N_\epsilon(\widetilde{\mathbb{X}})}. \end{aligned} \quad (28)$$

We opt for $t = O\left(\sqrt{\frac{\log N_\epsilon(\widetilde{\mathbb{X}})}{M}}\right)$, and subsequently, with probability $1 - e^{-\Omega(\log N_\epsilon(\widetilde{\mathbb{X}}))}$, we derive

$$\|\rho_{\text{PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{\log N_\epsilon(\widetilde{\mathbb{X}})}{M}}\right). \quad (29)$$

Proof of Theorem 3

Proof. We define a restricted Frobenius norm as following:

$$\begin{aligned} \|\rho_1 - \rho_2\|_{F,2r} &= \|\rho_1 - \rho_2\|_F \\ &= \max_{\rho \in \widetilde{\mathbb{X}}_{2r}} \langle \rho_1 - \rho_2, \rho \rangle, \end{aligned} \quad (30)$$

where the set $\widetilde{\mathbb{X}}_r$ is defined as follows:

$$\begin{aligned} \widetilde{\mathbb{X}}_r &= \left\{ \rho \in \mathbb{C}^{2^n \times 2^n} : \rho = \rho^\dagger, \right. \\ &\quad \left. \text{rank}(\rho) = r, \text{trace}(\rho) = 0, \|\rho\|_F \leq 1 \right\}. \end{aligned} \quad (31)$$

By the definition of the restricted Frobenius norm in Eq. (30), we can further analyze

$$\begin{aligned} \|\rho_{\text{LR-PCS}} - \rho^*\|_F &= \|\rho_{\text{LR-PCS}} - \rho^*\|_{F,2r} \\ &\leq \|\mathcal{P}_{\text{ED}}(\rho_{\text{CS}}) - \rho^*\|_{F,2r} \\ &\leq 2 \|\rho_{\text{CS}} - \rho^*\|_{F,2r} \\ &= 2 \max_{\rho \in \widetilde{\mathbb{X}}_{2r}} \left\langle \frac{1}{M} \sum_{m=1}^M [(2^n + 1) \phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}] - \rho^*, \rho \right\rangle, \end{aligned} \quad (32)$$

where the first two inequalities respectively follow the nonexpansiveness property of the projection and the quasi-optimality property of eigenvalue decomposition (ED) projection⁴². Next, we need to bound the first term in the last line of Eq. (32) using the covering argument. According to [ref. 59, Lemma 3.1], we initially construct an ϵ -net $\{\rho^{(1)}, \dots, \rho^{N_\epsilon(\tilde{\mathbb{X}}_{2r})}\} \in \tilde{\mathbb{X}}_{2r} \subset \hat{\mathbb{X}}_{2r}$ in which the size of $\tilde{\mathbb{X}}_{2r}$ is denoted by $N_\epsilon(\tilde{\mathbb{X}}_{2r}) \leq (\frac{9}{\epsilon})^{(2^{n+2}+2)r}$ such that

$$\sup_{\rho: \|\rho\|_F \leq 1} \min_{p \in N_\epsilon(\tilde{\mathbb{X}}_{2r})} \|\rho - \rho^{(p)}\|_F \leq \epsilon. \quad (33)$$

Combining Eqs. (23), (27), there exists an ϵ -net $\tilde{\mathbb{X}}_{2r}$ of $\hat{\mathbb{X}}_{2r}$ such that

$$\begin{aligned} & \mathbb{P} \left(\max_{\rho \in \tilde{\mathbb{X}}_{2r}} \left\langle \frac{1}{M} \sum_{m=1}^M [(2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}] - \rho^*, \rho \right\rangle \geq t \right) \\ & \leq \mathbb{P} \left(\max_{\rho^{(p)} \in \tilde{\mathbb{X}}_{2r}} \frac{1}{M} \sum_{m=1}^M \langle (2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^{(p)}, \rho^{(p)} \rangle \geq \frac{t}{2} \right) \\ & \leq x \mathbb{P} \left(\max_{\rho^{(p)} \in \tilde{\mathbb{X}}_{2r}} \frac{1}{M} \sum_{m=1}^M \langle (2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^{(p)}, \rho^{(p)} \rangle \geq \frac{t}{2} \right) \\ & \leq 2 \left(\frac{9}{\epsilon} \right)^{(2^{n+2}+2)r} e^{-\frac{Mt^2}{112}} \\ & \leq e^{-\frac{Mt^2}{112} + C2^nr}, \end{aligned} \quad (34)$$

where we set $\epsilon = \frac{1}{2}$ and C is a positive constant. We opt for $t = O\left(\sqrt{\frac{2^nr}{M}}\right)$ and subsequently, with probability $1 - e^{-\Omega(2^nr)}$, derive

$$\|\rho_{\text{LR-PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{2^nr}{M}}\right). \quad (35)$$

Proof of Theorem 4

Proof. We define a restricted Frobenius norm as follows:

$$\begin{aligned} \|\rho_1 - \rho_2\|_{F,2D} &= \|\rho_1 - \rho_2\|_F \\ &= \max_{\rho \in \hat{\mathbb{X}}_{2D}} \langle \rho_1 - \rho_2, \rho \rangle. \end{aligned} \quad (36)$$

where we denote by $\hat{\mathbb{X}}_D$ the normalized set of MPOs with bond dimension D :

$$\begin{aligned} \hat{\mathbb{X}}_D &= \left\{ \rho \in \mathbb{C}^{2^n \times 2^n} : \rho = \rho^\dagger, \|\rho\|_F \leq 1, \text{trace}(\rho) = 0, \right. \\ & \quad \left. \text{bond dimension}(\rho) = D \right\}. \end{aligned} \quad (37)$$

Note that the presence of additional orthonormal structures arises from the fact that, according to ref. 60, any TT form is equivalent to a left-orthogonal TT form⁴².

We define $\mathcal{P}_{\text{trace}}(\cdot)$ as a projection onto a convex set $\{\rho \in \mathbb{C}^{2^n \times 2^n} : \text{trace}(\rho) = 1\}$. By the definition of the restricted Frobenius

norm (36), we can derive

$$\begin{aligned} & \|\rho_{\text{MPO-PCS}} - \rho^*\|_F \\ & \leq \|\mathcal{P}_{\text{trace}}(\text{SVD}_D^{\text{tt}}(\rho_{\text{CS}})) - \rho^*\|_F \\ & = \|\mathcal{P}_{\text{trace}}(\text{SVD}_D^{\text{tt}}(\rho_{\text{CS}})) - \rho^*\|_{F,2D} \\ & \leq \|\text{SVD}_D^{\text{tt}}(\rho_{\text{CS}}) - \rho^*\|_{F,2D} \\ & \leq (1 + \sqrt{n-1}) \|\rho_{\text{CS}} - \rho^*\|_{F,2D} \\ & = (1 + \sqrt{n-1}) \max_{\rho \in \tilde{\mathbb{X}}_{2D}} \left\langle \frac{1}{M} \sum_{m=1}^M ((2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}) - \rho^*, \rho \right\rangle \\ & = (1 + \sqrt{n-1}) \max_{\rho \in \tilde{\mathbb{X}}_{2D}} \left\langle \frac{1}{M} \sum_{m=1}^M ((2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}) - \rho^*, \rho \right\rangle \end{aligned} \quad (38)$$

where the first two inequalities respectively follow from the nonexpansiveness property of the projection onto the convex set, while the third inequality is a consequence of the quasi-optimality property of TT-SVD projection⁴². Additionally, we denote

$$\begin{aligned} \hat{\mathbb{X}}_D &= \left\{ \rho \in \mathbb{C}^{2^n \times 2^n} : \rho = \rho^\dagger, \text{trace}(\rho) = 0, \right. \\ & \quad \rho(i_1 \dots i_n, j_1 \dots j_n) = X_1^{i_1 j_1} X_2^{i_2 j_2} \dots X_n^{i_n j_n}, \\ & \quad X_1^{i_1 j_1} \in \mathbb{C}^{1 \times D}, X_n^{i_n j_n} \in \mathbb{C}^{D \times 1}, X_\ell^{i_\ell j_\ell} \in \mathbb{C}^{D \times D}, \\ & \quad \left. \|\mathbf{L}(X_\ell)\| \leq 1, \ell \in [n-1], \|\mathbf{L}(X_n)\|_F \leq 1 \right\}. \end{aligned} \quad (39)$$

Based on $\|\rho\|_F = \|\mathbf{L}(X_n)\|_F \leq 1$ for a left-orthogonal TT form using [ref. 61, Eq.(44)], we obtain the last line.

Next, we will apply the covering argument to bound (38). For any fixed value of $\tilde{\rho} \in \tilde{\mathbb{X}}_{2D} \subset \hat{\mathbb{X}}_{2D}$, using Eq. (23), concentration inequality in Eq. (27) and Lemma 3, there exists an ϵ -net $\tilde{\mathbb{X}}_{2D}$ of $\hat{\mathbb{X}}_{2D}$ such that

$$\begin{aligned} & \mathbb{P} \left(\max_{\rho \in \tilde{\mathbb{X}}_{2D}} \left\langle \frac{1}{M} \sum_{m=1}^M ((2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n}) - \rho^*, \rho \right\rangle \geq t \right) \\ & \leq \mathbb{P} \left(\max_{\tilde{\rho} \in \tilde{\mathbb{X}}_{2D}} \frac{1}{M} \sum_{m=1}^M \langle (2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \tilde{\rho} \rangle \geq \frac{t}{2} \right) \\ & \leq \mathbb{P} \left(\max_{\tilde{\rho} \in \tilde{\mathbb{X}}_{2D}} \frac{1}{M} \sum_{m=1}^M \langle (2^n + 1)\phi_{m,j_m} \phi_{m,j_m}^\dagger - \mathbf{I}_{2^n} - \rho^*, \tilde{\rho} \rangle \geq \frac{t}{2} \right) \\ & \leq 2 \left(\frac{4n+6}{\epsilon} \right)^{4nD^2} e^{-\frac{Mt^2}{112}} \\ & \leq e^{-\frac{Mt^2}{112} + CnD^2 \log n}, \end{aligned} \quad (40)$$

where we set $\epsilon = \frac{1}{2}$ and C is a positive constant. We opt for $t = O\left(\sqrt{\frac{nD^2 \log n}{M}}\right)$ and subsequently, with probability $1 - e^{-\Omega(nD^2 \log n)}$, derive

$$\|\rho_{\text{MPO-PCS}} - \rho^*\|_F \leq O\left(\sqrt{\frac{n^2 D^2 \log n}{M}}\right). \quad (41)$$

Maximum Likelihood Estimation for Low-rank States and MPO states Maximum likelihood estimation (MLE) is a widely used technique for quantum state reconstruction. Under single-shot measurements, the MLE loss function can be formulated as follows^{47,62–66}:

$$\min_{\substack{\rho \geq 0, \\ \text{trace}(\rho)=1}} f(\rho) = -\frac{1}{M} \sum_{m=1}^M \log(\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, \rho \rangle). \quad (42)$$

However, the objective function in (42) does not leverage the structural properties inherent in quantum states. To address this limitation, we propose two MLE methods tailored for (1) low-rank states and (2) MPO states.

Low-rank MLE. When the density matrix is low-rank, we adopt a Riemannian gradient descent (RGD) algorithm on the unit Frobenius norm sphere. Specifically, for a quantum state $\rho^* \in \mathbb{C}^{2^n \times 2^n}$ satisfying $\text{trace}(\rho^*) = 1$ and $\rho^* \geq 0$, we can factorize it as $\rho^* = F^* F^{*\dagger}$, $F^* \in \mathbb{C}^{2^n \times r}$ with $\|F^*\|_F = 1$. This leads to the reformulated MLE objective:

$$\min_{\substack{F \in \mathbb{C}^{2^n \times r} \\ \|F\|_F = 1}} f_1(F) = -\frac{1}{M} \sum_{m=1}^M \log(\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, F F^\dagger \rangle).$$

The corresponding Riemannian gradient descent update reads:

$$\hat{F}_t = F_{t-1} - \mu \mathcal{P}_{T_F \text{Sp}}(\nabla_{F_1} f_1(F_{t-1})) \text{ and } F_t = \frac{\hat{F}_t}{\|\hat{F}_t\|_F},$$

where the Euclidean gradient is $\nabla_{F_1} f_1(F) = -\frac{1}{M} \sum_{m=1}^M \frac{\phi_{m,j_m} \phi_{m,j_m}^\dagger}{\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, F F^\dagger \rangle} F$ and $\mathcal{P}_{T_F \text{Sp}}(V) = V - \langle F, V \rangle F$ denotes the projection onto the tangent space $T_F \text{Sp} = \{F: \|F\|_F = 1\}$. Here, μ is the step size.

MPO-based MLE. When the density matrix admits an MPO representation with bond dimension D , we consider the constrained optimization problem:

$$\min_{\rho \in \mathbb{X}_D} f_2(\rho) = -\frac{1}{M} \sum_{m=1}^M \log(\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, \rho \rangle).$$

We solve (43) using a projected gradient descent (PGD) scheme:

$$\rho_t = \mathcal{P}_{\text{Simplex}}(\text{SVD}_D^{\text{tr}}(\rho_{t-1} - \mu \nabla_{\rho} f_2(\rho_{t-1}))),$$

where $\nabla_{\rho} f_2(\rho) = -\frac{1}{M} \sum_{m=1}^M \frac{\phi_{m,j_m} \phi_{m,j_m}^\dagger}{\langle \phi_{m,j_m} \phi_{m,j_m}^\dagger, \rho \rangle}$, and μ is the step size.

Materials

Lemma 1. (Classical Bernstein's inequality²³, Theorem 6) Let $s_1, \dots, s_n \in \mathbb{R}$ denote i.i.d. copies of a mean-zero random variable s that obeys $\mathbb{E}[|s|^p] \leq p! R^{p-2} \sigma^2 / 2$ for all integers $p \geq 2$, where $R, \sigma^2 > 0$ are constants. Then, for all $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n s_i\right| \geq t\right) \leq 2e^{-\frac{t^2/2}{n\sigma^2 + Rt}}. \quad (43)$$

Lemma 2. (ref. 29, Lemma 10) For any $A_i, A_i^* \in \mathbb{R}^{r_{i-1} \times r_i}, i \in \{1, \dots, N\}$, we have

$$\begin{aligned} A_1 A_2 \cdots A_N - A_1^* A_2^* \cdots A_N^* \\ = \sum_{i=1}^N A_1^* \cdots A_{i-1}^* (A_i - A_i^*) A_{i+1} \cdots A_N. \end{aligned} \quad (44)$$

Lemma 3. There exists an ϵ -net $\tilde{\mathbb{X}}_D$ for $\hat{\mathbb{X}}_D$ in Eq. (39) under the Frobenius norm, i.e., $\|\rho - \rho^{(p)}\|_F \leq \epsilon$ for $\rho^{(p)} \in \tilde{\mathbb{X}}_D$, obeying

$$N_\epsilon(\tilde{\mathbb{X}}_D) \leq \left(\frac{4n + \epsilon}{\epsilon}\right)^{4nD^2}, \quad (45)$$

where $N_\epsilon(\tilde{\mathbb{X}}_D)$ denotes the number of elements in the set $\tilde{\mathbb{X}}_D$.

Proof. For each set of matrices $\{L(X_\ell) \in \mathbb{R}^{4D \times D} : \|L(X_\ell)\| \leq 1\}$, according to ref. 67, we can construct an ξ -net $\{L(X_\ell^{(1)}), \dots, L(X_\ell^{(N_\ell)})\}$ with the covering number $N_\ell \leq \left(\frac{4+\xi}{\xi}\right)^{4D^2}$ such that

$$\sup_{L(X_\ell): \|L(X_\ell)\| \leq 1} \min_{p_\ell \leq N_\ell} \|L(X_\ell) - L(X_\ell^{(p_\ell)})\| \leq \xi, \quad (46)$$

for all $\ell \in \{1, \dots, n-1\}$. Also, we can construct an ξ -net $\{L(X_n^{(1)}), \dots, L(X_n^{(N_n)})\}$ for $\{L(X_n) \in \mathbb{R}^{4D \times 1} : \|L(X_n)\|_F \leq 1\}$ such that

$$\sup_{L(X_n): \|L(X_n)\|_F \leq 1} \min_{p_n \leq N_n} \|L(X_n) - L(X_n^{(p_n)})\|_F \leq \xi, \quad (47)$$

with the covering number $N_n \leq \left(\frac{2+\xi}{\xi}\right)^{4D}$.

Therefore, we can construct a ξ -net $\{[X_1^{(1)}, \dots, X_n^{(1)}], \dots, [X_1^{(N_1)}, \dots, X_n^{(N_n)}]\}$ with covering number

$$\prod_{\ell=1}^n N_\ell \leq \left(\frac{4+\xi}{\xi}\right)^{4nD^2} \quad (48)$$

for any MPO $\rho = [X_1, \dots, X_n]$ with bond dimension D . Then we expand $\|\rho - \rho^{(p)}\|_F$ as follows:

$$\begin{aligned} \|\rho - \rho^{(p)}\|_F &= \|[X_1, \dots, X_n] - [X_1^{(p_1)}, \dots, X_n^{(p_n)}]\|_F \\ &= \left\| \sum_{a_i=1}^n [X_1^{(p_1)}, \dots, X_{a_i-1}^{(p_{a_i})}, X_{a_i}^{(p_{a_i})} - X_{a_i}, X_{a_i+1}, \dots, X_n] \right\|_F \\ &\leq \sum_{a_i=1}^n \| [X_1^{(p_1)}, \dots, X_{a_i-1}^{(p_{a_i})}, X_{a_i}^{(p_{a_i})} - X_{a_i}, X_{a_i+1}, \dots, X_n] \|_F \\ &\leq \sum_{a_i=1}^{n-1} \|L(X_{a_i}^{(p_{a_i})}) - L(X_{a_i})\| + \|L(X_n^{(p_n)}) - L(X_n)\|_F \\ &\leq n\xi = \epsilon, \end{aligned}$$

where the second line and the second inequality respectively follow Lemma 2 and²⁹, Eq.(47). In addition, we choose $\xi = \frac{\epsilon}{n}$ in the last line. Ultimately, we can construct an ϵ -net $\{\rho^{(1)}, \dots, \rho^{(N_1 \cdots N_n)}\}$ with covering number

$$N_\epsilon(\tilde{\mathbb{X}}_D) \leq \left(\frac{4n + \epsilon}{\epsilon}\right)^{4nD^2} \quad (49)$$

for any MPO $\rho \in \hat{\mathbb{X}}_D$.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The code used in this study is available from the corresponding author upon reasonable request.

Received: 6 January 2025; Accepted: 15 August 2025;

Published online: 02 September 2025

References

- Bertrand, J. & Bertrand, P. A tomographic approach to Wigner's function. *Found. Phys.* **17**, 397–405 (1987).
- Vogel, K. & Risken, H. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A* **40**, 2847 (1989).
- Leonhardt, U. Quantum-state tomography and discrete wigner function. *Phys. Rev. Lett.* **74**, 4101 (1995).
- Hradil, Z. Quantum-state estimation. *Phys. Rev. A* **55**, R1561 (1997).

5. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
6. Haah, J., Harrow, A., Ji, Z., Wu, X. & Yu, N. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Theory* **63**, 5628–5641 (2017).
7. Fano, U. Description of states in quantum mechanics by density matrix and operator techniques. *Rev. Mod. Phys.* **29**, 74 (1957).
8. Reháček, J., Hradil, Z. & Ježek, M. Iterative algorithm for reconstruction of entangled states. *Phys. Rev. A* **63**, 040303 (2001).
9. Blume-Kohout, R. Optimal, reliable estimation of quantum states. *N. J. Phys.* **12**, 043034 (2010).
10. Granade, C., Combes, J. & Cory, D. Practical Bayesian tomography. *N. J. Phys.* **18**, 033024 (2016).
11. Lukens, J. M., Law, K. J. H., Jasra, A. & Lougovski, P. A practical and efficient approach for Bayesian quantum state estimation. *N. J. Phys.* **22**, 063038 (2020).
12. Blume-Kohout, R. Robust error bars for quantum tomography. *arXiv:1202.5270* (2012).
13. Faist, P. & Renner, R. Practical and reliable error bars in quantum tomography. *Phys. Rev. Lett.* **117**, 010404 (2016).
14. 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).
15. Kyriallidis, A. et al. Provable compressed sensing quantum state tomography via non-convex methods. *npj Quantum Inf.* **4**, 36 (2018).
16. Brandão, F. G., Kueng, R. & França, D. S. Fast and robust quantum state tomography from few basis measurements. *arXiv:2009.08216* (2020).
17. Zhu, Z., Lukens, J. M. & Kirby, B. T. On the connection between least squares, regularization, and classical shadows. *Quantum* **8**, 1455 (2024).
18. Sen, P., Bhatia, A. S., Bhangu, K. S. & Elbeltagi, A. Variational quantum classifiers through the lens of the Hessian. *Plos one* **17**, e0262346 (2022).
19. Liu, Y. et al. Variational quantum circuits for quantum state tomography. *Phys. Rev. A* **101**, 052316 (2020).
20. Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum principal component analysis. *Nat. Phys.* **10**, 631–633 (2014).
21. Kurmapu, M. K. Machine learning assisted quantum state tomography (2020).
22. Kueng, R., Rauhut, H. & Terstiege, U. Low rank matrix recovery from rank one measurements. *Appl. Comput. Harmon. Anal.* **41**, 88–116 (2017).
23. Guță, M., Kahn, J., Kueng, R. & Tropp, J. A. Fast state tomography with optimal error bounds. *J. Phys. A: Math. Theor.* **53**, 204001 (2020).
24. Voroninski, V. Quantum tomography from few full-rank observables. *arXiv:1309.7669* (2013).
25. Liu, Y.-K. Universal low-rank matrix recovery from Pauli measurements. *Adv. Neural Inf. Process Syst* **24**, 1–9 (2011).
26. Eisert, J., Cramer, M. & Plenio, M. B. Colloquium: Area laws for the entanglement entropy. *Rev. Mod. Phys.* **82**, 277–306 (2010).
27. Pirvu, B., Murg, V., Cirac, J. I. & Verstraete, F. Matrix product operator representations. *N. J. Phys.* **12**, 025012 (2010).
28. Noh, K., Jiang, L. & Fefferman, B. Efficient classical simulation of noisy random quantum circuits in one dimension. *Quantum* **4**, 318 (2020).
29. Qin, Z., Jameson, C., Gong, Z., Wakin, M. B. & Zhu, Z. Quantum state tomography for matrix product density operators. *IEEE Trans. Inf. Theory* **70**, 5030–5056 (2024).
30. Qin, Z. et al. Sample-optimal quantum state tomography for structured quantum states in one dimension. *arXiv:2410.02583* (2024).
31. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
32. Acharya, A., Saha, S. & Sengupta, A. M. Shadow tomography based on informationally complete positive operator-valued measure. *Phys. Rev. A* **104**, 052418 (2021).
33. Akhtar, A. A., Hu, H.-Y. & You, Y.-Z. Scalable and flexible classical shadow tomography with tensor networks. *Quantum* **7**, 1026 (2023).
34. Grier, D., Pashayan, H. & Schaeffer, L. Sample-optimal classical shadows for pure states. *Quantum* **8**, 1373 (2024).
35. Ippoliti, M. Classical shadows based on locally-entangled measurements. *Quantum* **8**, 1293 (2024).
36. Becker, S., Datta, N., Lami, L. & Rouz, C. Classical shadow tomography for continuous variables quantum systems. *IEEE Trans. Inf. Theory* **70**, 3427–3452 (2024).
37. Huang, H.-Y., Preskill, J. & Soleimanifar, M. Certifying almost all quantum states with few single-qubit measurements. 1202–1206 (IEEE, 2024).
38. Wei, V., Coish, W., Ronagh, P. & Muschik, C. A. Neural-shadow quantum state tomography. *Phys. Rev. Res.* **6**, 023250 (2024).
39. Struchalin, G., Zagorovskii, Y. A., Kovlakov, E., Straupe, S. & Kulik, S. Experimental estimation of quantum state properties from classical shadows. *PRX Quantum* **2**, 010307 (2021).
40. Kokaew, W., Kulchytsky, B., Matsuura, S. & Ronagh, P. Bootstrapping classical shadows for neural quantum state tomography. *arXiv:2405.06864* (2024).
41. Chen, Y. & Ye, X. Projection onto a simplex. *arXiv:1101.6081* (2011).
42. Oseledets, I. Tensor-train decomposition. *SIAM J. Sci. Comput.* **33**, 2295–2317 (2011).
43. Nielsen, M. A. & Chuang, I. L. Quantum computation and quantum information. (Cambridge university press, 2000).
44. Schwemmer, C. et al. Systematic errors in current quantum state tomography tools. *Phys. Rev. Lett.* **114**, 080403 (2015).
45. 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).
46. Lvovsky, A. I. Iterative maximum-likelihood reconstruction in quantum homodyne tomography. *J. Opt. B: Quantum Semiclass. Opt.* **6**, S556 (2004).
47. Baumgratz, T., Nüeler, A., Cramer, M. & Plenio, M. B. A scalable maximum likelihood method for quantum state tomography. *N. J. Phys.* **15**, 125004 (2013).
48. Jameson, C. et al. Optimal quantum state tomography with local informationally complete measurements. *arXiv:2408.07115* (2024).
49. Lukens, J. M., Law, K. J. & Bannink, R. S. A Bayesian analysis of classical shadows. *npj Quantum Inf.* **7**, 113 (2021).
50. Hu, H.-Y. & You, Y.-Z. Hamiltonian-driven shadow tomography of quantum states. *Phys. Rev. Res.* **4**, 013054 (2022).
51. Hu, H.-Y., Choi, S. & You, Y.-Z. Classical shadow tomography with locally scrambled quantum dynamics. *Phys. Rev. Res.* **5**, 023027 (2023).
52. Candes, E. J. & Plan, Y. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inf. Theory* **57**, 2342–2359 (2011).
53. Werner, A. H. et al. Positive tensor network approach for simulating open quantum many-body systems. *Phys. Rev. Lett.* **116**, 237201 (2016).
54. Cramer, M. et al. Efficient quantum state tomography. *Nat. Commun.* **1**, 1–7 (2010).
55. Cirac, J. I., Perez-Garcia, D., Schuch, N. & Verstraete, F. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Rev. Mod. Phys.* **93**, 045003 (2021).
56. Haegeman, J., Osborne, T. J., Verschelde, H. & Verstraete, F. Entanglement renormalization for quantum fields in real space. *Phys. Rev. Lett.* **110**, 100402 (2013).
57. Mele, A. A. Introduction to Haar measure tools in quantum information: A beginner's tutorial. *arXiv:2307.08956* (2023).

58. Zhu, Z., Li, Q., Tang, G. & Wakin, M. B. The global optimization geometry of low-rank matrix optimization. *IEEE Trans. Inf. Theory* **67**, 1308–1331 (2021).
59. Candès, E. J., Li, X., Ma, Y. & Wright, J. Robust principal component analysis? *J. ACM* **58**, 1–37 (2011).
60. Holtz, S., Rohwedder, T. & Schneider, R. On manifolds of tensors of fixed tt-rank. *Numer. Math.* **120**, 701–731 (2012).
61. Qin, Z., Wakin, M. B. & Zhu, Z. Guaranteed nonconvex factorization approach for tensor train recovery. *arXiv:2401.02592* (2024).
62. Hradil, Z., Řeháček, J., Fiurášek, J. & Ježek, M. 3 maximum-likelihood methods in quantum mechanics. *Quantum State Estimation* 59–112 (2004).
63. Gonçalves, D. S., Gomes-Ruggiero, M. A., Lavor, C., Farias, O. J. & Ribeiro, P. S. Local solutions of maximum likelihood estimation in quantum state tomography. *Quantum Inf. Comput.* **12**, 775–790 (2012).
64. Silva, G., Glancy, S. & Vasconcelos, H. M. Investigating bias in maximum-likelihood quantum-state tomography. *Phys. Rev. A* **95**, 022107 (2017).
65. Shang, J., Zhang, Z. & Ng, H. K. Superfast maximum-likelihood reconstruction for quantum tomography. *Phys. Rev. A* **95**, 062336 (2017).
66. Scholten, T. L. & Blume-Kohout, R. Behavior of the maximum likelihood in quantum state tomography. *N. J. Phys.* **20**, 023050 (2018).
67. Zhang, A. & Xia, D. Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64**, 7311–7338 (2018).

Acknowledgements

We acknowledge funding support from the National Science Foundation (CCF-2241298, EECS-2409701) and the U.S. Department of Energy (ERKJ432, DE-SC0024257). We thank the Ohio Supercomputer Center for providing the computational resources and the Quantum Collaborative led by Arizona State University for providing valuable expertise and resources. A portion of this work was performed at Oak Ridge National Laboratory, operated by UT-Battelle for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Author contributions

Z.Q. led the conceptualization, developed the methodology, implemented the software, and drafted the manuscript. J.M.L. and B.K. contributed to the conceptual design, methodological development, and manuscript revision. Z.Z. supervised the project and contributed to the conceptual design, methodological development, and manuscript revision.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Zhihui Zhu.

Reprints and permissions 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.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025