Ultrafast quantum state tomography with feed-forward neural networks

Yong Wang\(^1\), Shuming Cheng\(^{1,2,3}\), Li Li\(^{1,2}\), and Jie Chen\(^{1,2}\)

\(^1\)College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China
\(^2\)Shanghai Research Institute for Artificial Intelligence Science and Technology, Tongji University, Shanghai 201203, China
\(^3\)Institute for Advanced Study, Tongji University, Shanghai, 200092, China

Reconstructing the state of many-body quantum systems is of fundamental importance in quantum information tasks, but extremely challenging due to the curse of dimensionality. In this work, we present a quantum tomography approach based on neural networks to achieve the ultrafast reconstruction of multi-qubit states. Particularly, we propose a simple 3-layer feed-forward network to process the experimental data generated from measuring each qubit with a positive operator-valued measure, which is able to reduce the storage cost and computational complexity. Moreover, the techniques of state decomposition and \(P\)-order absolute projection are jointly introduced to ensure the positivity of state matrices learned in the maximum likelihood function and to improve the convergence speed and robustness of the above network. Finally, it is tested on a large number of states with a wide range of purity to show that we can faithfully tomography 11-qubit states on a laptop within 2 minutes under noise. Our numerical results also demonstrate that more state samples are required to achieve the given tomography fidelity for the low-purity states, and the increased depolarizing noise induces a linear decrease in the tomography fidelity.

1 Introduction

Quantum state tomography (QST) is a powerful tool to recover the state information of the unknown many-body quantum systems from measurement statistics, and hence plays an indispensable role in quantum information processing tasks, with wide applications ranging from certifying fundamental principles in quantum theory \cite{1}, benchmarking quantum devices \cite{2}, to verifying quantum algorithms \cite{3}. However, it is a challenging task due to the curse of dimensionality in the sense that it admits an exponential growth of measurement settings, memory cost, and computing resources as the number of subsystems involved linearly increases \cite{4}.

Various approaches have been developed to accomplish the task of QST over the past decades. Notable examples include the maximum likelihood estimation (MLE) \cite{5, 6, 7}, linear regression \cite{8, 9}, and Bayesian tomography \cite{10, 11}. The state reconstruction speed could be improved by adopting compressed sensing \cite{12, 13, 14}, permutationally invariant tomography \cite{15, 16}, or matrix-product-state tomography \cite{17, 18}, however, they are only applicable for limited classes of states and thus lack generality \cite{19}. Furthermore, since one fundamental physical constraint on the matrices reconstructed from these methods is the positivity, some state-mapping techniques are needed. For example, one is to parameterize the state matrix via the Cholesky decomposition \cite{5, 7, 10, 19, 13, 14, 20, 21} and another is the state projection which maps negative eigenvalues of the learned matrix to those of physical states \cite{8, 9, 22, 11, 23, 24}. However, the problematic issues arise that the decomposition method has limited convergence speed and reconstruction accuracy while the projection needs extra computing resources \cite{25}. Hence, it still remains rather challenging for these methods to tomography multi-dimensional quantum systems.

As the neural networks become more efficient and scalable to process the data of high dimensionality, many neural-network QST (NN-QST)
approaches have also been proposed, such as feed-forward neural network (FNN) [26, 27, 28, 20, 21], restricted Boltzmann machines [29, 30, 31], recurrent neural networks [32, 33], autoregressive models [34], generative adversarial networks (GANs) [19], and transformers [35]. It has been demonstrated in [19] that half an hour is enough to reconstruct the 6-qubit cat state by using conditional GANs. However, it is still less explored how the neuralnet works for multi-qubit states with a wide range of purity and whether the techniques of state decomposition and/or projection can be applied to these network-based QST to improve their learning performances.

To tackle the above issues, here we propose a feed-forward neural network, combining the state decomposition with state projection, to achieve the ultrafast reconstruction of multi-qubit states. In particular, our network is configured in a simple 3-layer structure where the input layer directly processes the probability distributions (PDs) generated by measuring each qubit, the hidden layer has fully-connected neurons in a linear number of qubits, and the output yields a state matrix directly. Then, it is efficiently trained by optimizing the maximum likelihood function and benchmarked by quantum state fidelity between the learned state and the target. In this training process, the transition matrix is decomposed into Hermitian matrices which are then processed via a class of the projection methods called as \(P\)-order absolute projection. Thus, we provide the first unified state-mapping strategy which is able to not only speed-up its learning rate, but also to improve its robustness towards state purity and noise. Finally, a positive operator-valued measure (POVM) on each qubit in the product structure is used to process the PDs, which reduces the memory cost to store and saves the computational resource to process.

We further test our NN-QST on a large number of states, including trivial product states, genuinely entangled W and GHZ\(i\) states, all of these states under depolarizing noise, and states randomly generated with exponential decay of eigenvalues and fixed state purity as introduced in [23]. Numerical results first confirm our NN-QST adopting the unified state-mapping strategy has better convergence speed and state purity robustness than those using other state-mapping methods. Next, we demonstrate that it has better performance in the convergence time and robustness towards state purity and noise, in comparison to various fast QST algorithms including the CG-APG algorithm [25] and iterative MLE (iMLE) algorithm [36]. Surprisingly, it is found our approach can fully tomography 11-qubit states on a laptop within 2 minutes, while it takes nearly 3 hours for the CG-APG and iMLE algorithms to accomplish the 10-qubit tomography and more than half an hour for conditional GANs to tomography the 6-qubit cat state [19]. Moreover, we show more state samples are needed for the net to achieve a given tomography fidelity for the low-purity states. Finally, we investigate how the state purity affects the tomography fidelity under noise and find it is robust to depolarizing noise.

The reminder of this work is organized as follows. Sec. 2 introduces some basic concepts related to QST, including the density matrix, MLE strategy, state-mapping techniques, and quantum state fidelity. Sec. 3 details the implementation of our NN-QST, unified state-mapping strategy, and the product-structured POVM. Then, the corresponding numerical results are presented in Sec. 4. Finally, the summary and outlook of the paper are discussed in Sec. 5.

2 Quantum state tomography

This section briefly introduces basic notations about QST and some commonly-used techniques, such as state decomposition and projection, which are used to guarantee the positivity of state matrices reconstructed from QST. Finally, quantum state fidelity is introduced to evaluate the tomography performance.

2.1 Reconstructing the density matrix

The state of many-body quantum systems is fully characterized by the density matrix \(\rho\), which is a positive semi-definite (PSD) operator with unit trace, i.e., \(\text{Tr}[\rho] = 1\). Specifically, it can be written in a form of

\[
\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.
\]  

(1)

where the probability \(p_i\) represents the occurrence of pure quantum state \(|\psi_i\rangle\), with \(\sum_i p_i = 1\) and \(p_i \geq 0\). If \(\rho = |\psi\rangle\langle\psi|\), then \(\rho\) is pure. Otherwise, it is a mixed state. Note that for the \(N\)-qubit
system, generically, it requires \( d^2 - 1 \) real parameters to completely determine \( \rho \), where \( d = 2^N \) is the dimension of the Hilbert space. It is evident that as the number of qubits \( N \) increases, the number of parameters to describe \( \rho \) will increase exponentially.

To reconstruct the density matrix \( \rho \) in Eq. (1), QST is decomposed of two steps: The first is the measurement procedure which yields the outcome statistics described by PDs from measuring on identically prepared copies of the unknown quantum state. Each measurement \( m \) is modeled as a positive-operator-value measure (POVM) \( \{ M_k \} \) where the positive operator \( M_k \) satisfy \( \sum_k M_k = I \) for all \( m \). And, the second is the estimation procedure which estimate a physical \( \hat{\rho} \) from those measured PDs. Consequently, it can be formulated as the following optimization problem:

\[
\hat{\rho} \quad \text{s.t.} \quad \text{Tr}[\hat{\rho}] = 1 \quad \text{and} \quad \hat{\rho} \geq 0, \\
P_k = \text{Tr}[M_k \hat{\rho}] = \text{Tr}[M_k \rho] \quad \forall \ M_k, \\
M_k \geq 0 \quad \text{and} \quad \sum_k M_k = I.
\]

However, it is challenging to directly solve the above problem due to the curse of dimensionality that as the number of qubits linearly increases, it requires an exponential growth of measurement settings to faithfully determine \( \rho \) memory cost to store \( P^i = (P_1^i, \ldots, P_n^i) \), and computational resources to process the PDs. In practice, the presence of statistical noise leads us to obtain the frequency \( f_k \propto \text{Tr}[M_k \rho] \), instead of the accurate probability of each outcome, which makes the accurate tomography more challenging.

One efficient approach to QST is the maximum likelihood estimation [5, 6, 7]. In particular, MLE minimizes the negative log-likelihood function between the observed frequency \( f_k \) and the estimated probability \( \hat{P}_k = \text{Tr}[M_k \hat{\rho}] \):

\[
\text{minimize} \quad - \sum_k f_k \log (\text{Tr}[M_k \rho]) \\
\text{s.t.} \quad \text{Tr}[\hat{\rho}] = 1 \quad \text{and} \quad \hat{\rho} \geq 0, \\
M_k \geq 0 \quad \text{and} \quad \sum_k M_k = I.
\]

where \( \{ M_k \} \) is one POVM, and \( \hat{\rho} \) is the reconstructed physical density matrix. Notably, it follows from the convexity of the optimization function that there exists an unique solution to the MLE [37, 38]. Since no extra assumption is imposed by the MLE, it is also one of commonly-used tomography methods [39], together with experimental verification via the single-ion Zeeman qubit [40] and the polarization states of three photons [41].

2.2 Mapping techniques to ensure the positivity of reconstructed states

In the optimization problems (2) and (3), the reconstructed state \( \hat{\rho} \) should satisfy the fundamental physical constraint that it is a PSD operator with trace being 1, however, this may not be always guaranteed. Thus, to solve this problem, the techniques of state decomposition or state projection are further employed to map the nonphysical states to physical ones.

2.2.1 Decomposition method

Note that an arbitrary density matrix \( \hat{\rho} \) admits the Cholesky decomposition,

\[
\hat{\rho} = \frac{T_\rho^T T_\rho}{\text{Tr}[T_\rho^T T_\rho]},
\]

where the transition matrix \( T_\rho \) is a complex lower triangular matrix and \( \dagger \) denotes the complex conjugate operation. Evidently, \( \hat{\rho} \) in the form (4) is automatically positive, and hence the QST optimization (2) and (3) can be simplified to search over \( T_\rho \) without the positivity constraint, instead of \( \hat{\rho} \).

It immediately yields that the MLE with the state decomposition (4) becomes:

\[
\text{minimize} \quad - \sum_k f_k \log (\text{Tr}[M_k \frac{T_\rho^T T_\rho}{\text{Tr}[T_\rho^T T_\rho]}]) \\
\text{s.t.} \quad M_k \geq 0 \quad \text{and} \quad \sum_k M_k = I.
\]

Here \( T_\rho \) could be a nonphysical state. It is remarked that the difficulty in ensuring the positive \( \hat{\rho} \) is essentially transferred to optimize the objective function in the above problems, and there is evidence to signal the slow convergence of these methods with decomposition [25]. Hence, another technique, called as state projection method, has been proposed to speed up the estimation process [8, 9, 11, 22, 23, 24].
2.2.2 Projection method

Given a nonpositive Hermitian matrix $\hat{\rho}$ reconstructed from QST, the projection method introduces a map $\mathcal{P}(\bullet)$ which maps its eigenvalues to nonnegative ones with unit sum to produce a physical matrix $\hat{\rho}$. Indeed, following from the eigenvalue decomposition $\hat{\rho} = Q\Lambda Q^\dagger$ where $\Lambda$ is the diagonal eigenvalue matrix and $Q$ is the corresponding orthogonal matrix, we have

$$\hat{\rho} = \mathcal{P}(\hat{\rho}) = \mathcal{P}(Q\Lambda Q^\dagger) = Q\mathcal{P}(\Lambda)Q^\dagger = Q\Sigma Q^\dagger. \quad (6)$$

Here $\Sigma = \mathcal{P}(\Lambda)$ is a positive diagonal matrix with trace being 1, thus satisfying the physical constraints imposed by density matrices. There are various ways to construct the map in Eq. (6), and one possible way is subtracting a coefficient responding orthogonal matrix, we have

$$\Sigma_i = \mathcal{P}(\Lambda_i) = \max(\Lambda_i, 0) \quad (7)$$

for each eigenvalue $\Lambda_i$. The known examples include the Frobenius norm with alternative projection $\mathcal{F}[]$ [22] and simplex projection $\mathcal{S}[]$ [23] to choose a proper coefficient $c$ or the nuclear norm with projection to set $c$ to 0 [11]. For the non-Hermitian matrix $\hat{\rho}$, using $(\hat{\rho} + \hat{\rho}^\dagger)/2$ naturally gives rise to a Hermitian matrix which then could be processed in the same procedure.

Then, the MLE (3) can be solved with the projection method

$$\text{minimize } -\sum_k f_k \log(\text{Tr} [M_k \mathcal{P}(\hat{\rho})])$$

s.t. $M_k \geq 0$ and $\sum_k M_k = 1. \quad (8)$$

where $\mathcal{P}(\bullet)$ is any chosen projection method. Essentially, the projection method pulls the estimate matrix $\hat{\rho}$ close to a physical density matrix in the sense of matrix 2-norm [22, 23] and other operator norms [11], to guarantee convergence speed and accuracy.

2.3 Quantum state fidelity

The QST performance can be evaluated via quantum state fidelity which measures the distance between quantum states. Specifically, the quantum fidelity $F_q$ for the state matrix $\hat{\rho}$ reconstructed from QST and the target $\rho$ is defined as [42]

$$F_q(\hat{\rho}, \rho) = \left(\text{Tr} \left[\sqrt{\sqrt{\rho} \hat{\rho} \sqrt{\rho}}\right]\right)^2. \quad (9)$$

The state fidelity is symmetric, i.e., $F_q(\hat{\rho}, \rho) = F_q(\rho, \hat{\rho})$. If the target state is pure, then Eq. (9) simplifies to $F_q(\rho = |\psi\rangle\langle\psi|, \hat{\rho}) = \langle\psi|\hat{\rho}|\psi\rangle$. Generally, computing Eq. (9) requires multiple square-root matrix operations, which consume large computational resources. However, this computing cost can be reduced, if the state information in QST is used.

Noting from the state projection method in Eq. (6) that $\hat{\rho} = \mathcal{P}(\hat{\rho}) = Q\Sigma Q^\dagger$, we have

$$F_q(\hat{\rho}, \rho) = \left(\text{Tr} \left[\sqrt{Q\sqrt{\Sigma} Q^\dagger}\rho \sqrt{Q\sqrt{\Sigma} Q^\dagger}\right]\right)^2$$

$$= \left(\text{Tr} \left[\sqrt{Q\Sigma Q^\dagger}\right]\right)^2$$

$$= \left(\text{Tr} \left[Q'\sqrt{\Sigma'} Q'^\dagger\right]\right)^2$$

$$= \left(\sum_i \sqrt{\Sigma_i'}\right)^2. \quad (10)$$

The second equality follows from the eigenvalue decomposition $Q\sqrt{\Sigma} Q^\dagger \rho Q\sqrt{\Sigma} Q^\dagger = Q\Sigma' Q'^\dagger \rho Q\Sigma' Q'^\dagger$ with the orthogonal matrix $Q'$ and nonnegative diagonal matrix $\Sigma'$. Obviously, the fidelity (10) needs one single matrix-decomposition operation, thus reducing much computational complexity.

Finally, quantum state fidelity $F_q$ can be upper bounded by the classical fidelity $F_c$ between two probability distributions, i.e.,

$$F_q(\hat{\rho}, \rho) \leq F_c(\hat{\rho}, \rho|\rho_k) = \left(\sum_k \sqrt{\hat{P}_k P_k}\right)^2, \quad (11)$$

where $\hat{P}_k = \text{Tr} [M_k \hat{\rho}]$ and $P_k = \text{Tr} [M_k \rho]$. It is easy to calculate the classical fidelity which provides useful bounds for benchmarking QST. Further, $F_c$ can be approximated by sampling

$$F_c(\hat{\rho}, \rho|M_k) = \left(\sum_k \hat{P}_k \sqrt{\frac{P_k}{P_k}}\right)^2 \approx \left(\mathbb{E}_{k \sim \hat{\rho}} \left[\frac{\hat{P}_k}{P_k}\right]\right)^2. \quad (12)$$

The sampled value is generally smaller than the true value $F_c$.

3 Ultrafast neural-network quantum state tomography

As mentioned above, it becomes much less efficient for the traditional QST approaches with the techniques of state decomposition and/or projection to tomography the multi-qubit system with...
a large number of qubits. In this section, we propose a QST approach which employs the neural network to speed up the data-processing process and hence to achieve the ultrafast state reconstruction. Surprisingly, it is found that we are able to achieve the full tomography of 11-qubit states within 2 minutes, while it takes nearly 3 hours to tomography the 10-qubit state for the CG-APG algorithm [25] and iMLE algorithm [36]. The framework of our ultrafast NN-QST is shown in Fig. 1. First, the network is built up as a simple 3-layer feed-forward neural network (FNN) which processes the experimental data encoded by PDs and output a Hermitian matrix $\hat{T}_\rho$. Then, the state decomposition method is used to ensure the positive $\hat{\rho}$ and the projection method is further introduced to improve the convergence speed of the training process. Moreover, the product-structured POVM is employed to store and process the data, which can reduce the memory cost and computational complexity. Finally, the reconstructed state matrix is benchmarked via quantum state fidelity.

3.1 The ultrafast feed-forward neural network

The FNN used in this work is easy to implement as it is only composed of the input, hidden, and output layers. For the $N$-qubit system, the number of neurons in each layer is chosen as:

1. Input layer: the maximum is $N_{\text{in}} = K^N$ equal to the dimension of the PDs generated from measuring a POVM with $K$ elements on each qubit.

2. Output layer: $N_{\text{out}} = 4^N$ which is the number of real parameters to determine the density matrix $\rho$ for a $N$-qubit state.

3. Hidden layer: $N_{\text{hidden}} = \log N_{\text{out}} = 2N$ which is a linear number of qubits.

The loss function for the network is chosen as the negative log-likelihood function

$$L(\hat{\rho}) = -\sum_k f_k \log(\text{Tr}[M_k \hat{\rho}]), \quad (13)$$

which is commonly used in maximum likelihood tomography. Here $\hat{\rho}$ refers to the learned density matrix, $f_k$ is the measurement frequency, and $M_k$ is one POVM element. And its performance in reconstructing the target state $\rho$ is evaluated via quantum state fidelity $F_\rho(\hat{\rho}, \rho)$ in (9).

Noting that other neural nets can also be used to accomplish the QST task, we propose the architecture of convolutional neural networks (CNNs) for QST in Appendix A. We find that the CNN based QST requires preprocessing the PDs and the convolution operations consume more computational resources, hence leading to a slower convergence speed than that of the FNN used here, when the number of qubits is greater than 8. A detailed experimental comparison between the FNN and CNN is also given in the Appendix A.

3.2 The unified state-mapping strategy

We present a simple and efficient method which combines the state decomposition with state projection to map the output matrix $T_{\hat{\rho}}$ to a physical state $\hat{\rho}$. Note first that if the matrix $T_{\hat{\rho}}$ is Hermi-
tian, then there is
\[
\hat{\rho} = \frac{T_\rho^\dagger T_\rho}{\text{Tr}[T_\rho^\dagger T_\rho]} = \frac{Q\Lambda Q^\dagger Q\Lambda Q^\dagger}{\text{Tr}[Q\Lambda Q^\dagger Q\Lambda Q^\dagger]}
\]
\[
= \frac{Q\Lambda^2 Q^\dagger}{\text{Tr}[\Lambda^2]}
\]
\[
= Q\Sigma Q^\dagger.
\]

Here \(T_\rho\) admits the eigenvalue decomposition \(Q\Lambda Q^\dagger\), and \(\Sigma = \Lambda^2/\text{Tr}[\Lambda^2]\) is automatically a nonnegative diagonal matrix with unit trace. This can be regarded as a unified strategy which first uses a simple projection method to process the transition matrix \(T_\rho\) and simultaneously use the state decomposition method to output a physical matrix \(\hat{\rho}\) from the processed \(T_\rho\).

Following then from the state projection (6), we are able to easily generalize the above result to
\[
\hat{\rho} = \frac{\mathcal{P}(T_\rho^\dagger)\mathcal{P}(T_\rho)}{\text{Tr}[\mathcal{P}(T_\rho^\dagger)\mathcal{P}(T_\rho)]} = \frac{QP(\Lambda)^2 Q^\dagger}{\text{Tr}[P(\Lambda)^2]}.
\]

Here we introduce a new class of projection methods: \(P\)-order absolute map defined as
\[
\mathcal{A}(T_\rho)_P = Q|\Lambda|^{P/2}Q^\dagger,
\]
where the tunable parameter \(P\) can be used to adjust the weight of different eigenvalues of \(\Lambda\), and it immediately gives rise to \(P\)-order absolute projection
\[
\hat{\rho} = \mathcal{A}[\cdot]_P = \frac{Q|\Lambda|^{P/2}Q^\dagger}{\text{Tr}[|\Lambda|^{P}]}.
\]

It is easy to find Eq. (14) is a special case of \(\mathcal{A}[\cdot]_P\) with \(P = 2\). Since more parameter degrees of freedom is assigned to the above projection (17), it is suitable for dealing with more complex tomographic conditions, which is numerically confirmed in the following sections.

Finally, it is remarked that there exists an alternate way to combine these two methods in which the decomposition method is used in the early stage to achieve a fast convergence speed and then the projection method is adopted in the latter stage to guarantee a high learning accuracy [25]. However, our numerical results in Sec. 4.4 show that it becomes inefficient to tomography the large-qubit state and its implementation in neural networks is extremely difficult due to switching. Besides, the issues about the switching conditions between these two methods and the corresponding computational cost remain to be examined carefully.

3.3 The product-structured POVM

Suppose a POVM \(\{M_k\}_{k=1}^K\) is performed on each qubit. Then, the product structure of single-qubit POVM is used to form a general general POVM \(\{M_k\}\) for the \(N\)-qubit state where \(k = (k_1, \ldots, k_N)\), with \(k_i \in \{1, \ldots, K\}\) and \(M_k = M_{k_1} \otimes \cdots \otimes M_{k_N}\). The number of measurement elements \(\{M_k\}\) is \(K^N\), thus yielding a \(K^N\)-dimensional PD.

For any \(N\)-qubit POVM in the above product structure, the tensor product and trace operations for obtaining \(P_k = \text{Tr}[M_k\rho]\) can be converted into the product of matrices. We take the 2-qubit case as an example. Given matrices
\[
M_i = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \quad \text{and} \quad \rho = \begin{bmatrix} A & B \\ C & D \end{bmatrix},
\]
where \(A, B, C, D\) are block matrix elements, there is
\[
P_{(i,j)} = \text{Tr}[(M_i \otimes M_j)\rho]
\]
\[
= \text{Tr}[(M_i \otimes M_j)\left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes A + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes B + \cdots \right)]
\]
\[
= a_i \text{Tr}[M_j A] + b_i \text{Tr}[M_j B] + c_i \text{Tr}[M_j C] + d_i \text{Tr}[M_j D]
\]
\[
= \mathcal{V}_{M_i}^T \mathcal{V}_A \mathcal{V}_B \mathcal{V}_C \mathcal{V}_D \mathcal{V}_{M_j}.
\]

And \(\mathcal{V}_X\) is the column vectorization of matrix \(X\), such as \(\mathcal{V}_{M_i}^T = \begin{bmatrix} a & b & c & d \end{bmatrix}\). This form can greatly reduce the computational and storage costs, which has been verified in [25] the cost of computing the probabilities is reduced from \(O(K^N 4^N)\) to \(O(K^{N+1})\). Thus, we choose the product-structured POVM in the FNN.

4 Numerical experiments and results

We test our NN-QST on a large number of multi-qubit states with a wide range of purity in this section. Specifically, the target states are chosen as:

- the \(N\)-qubit states
\[
\rho = p|\psi\rangle\langle\psi| + \frac{1-p}{d}\mathbb{I} \quad (19)
\]
with $0 \leq p \leq 1$ and $d = 2^N$. And the pure state $|\psi\rangle$ ranges from the product state

$$|\psi\rangle_p = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^\otimes N,$$

(20)

to genuinely entangled GHZi and W states respectively, i.e.,

$$|\psi\rangle_G = \left(\frac{|0\rangle^\otimes N + i|1\rangle^\otimes N}{\sqrt{2}}\right),$$

(21)

$$|\psi\rangle_W = \frac{1}{\sqrt{N}}(|00\ldots\rangle + \cdots + |00\ldots\rangle).$$

(22)

- the $N$-qubit states randomly generated with exponential decay of eigenvalues and fixed state purity [23].
- the $N$-qubit pure state $|\psi\rangle$ passing through the depolarizing channel

$$\rho = (1 - \lambda)|\psi\rangle\langle\psi| + \frac{\lambda}{d}$$

(23)

where the noise strength $0 \leq \lambda \leq 1$, and the pure state $|\psi\rangle$ from the product, GHZi, and W states. Note that although this class is equivalent to Eq. (19), the target states used to benchmark the QST methods are distinct in the sense that the former are pure states while the latter chooses states after the depolarizing channel.

Numerical experiments are then performed to study the following problems:

P1. How different state decomposition and/or projection methods affect the performance of the NN-QST.

P2. How the state purity affects the tomography fidelity of the NN-QST using different state-mapping methods.

P3. How much advantage in the convergence time does the NN-QST provide, in comparison to previous QST algorithms.

P4. How robust towards statistical and depolarizing noise are the NN-QST and other QST algorithms.

4.1 Experimental setups

Key settings in the FNN are given below:

- The Leaky rectified linear units activation function is used in the hidden layer.
- The Rprop optimizer is used and the learning rate is fixed at 0.001.

And the single-qubit POVM is chosen as

$$\{M_a = \frac{1}{4} (\mathbb{I} + s_a \cdot \sigma)\}_{a \in \{0, 1, 2, 3\}}$$

with

$$s_0 = \left(\frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}}\right), s_1 = \left(-\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}}\right),$$

$$s_2 = \left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}}\right), s_3 = \left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}}\right)$$

(24)

where $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli operator vectors. Evidently, we have $K = 4$. Note that our method can be easily applied to the case involving multiple POVMs on each qubit and adaptive local measurements. Here, in order to save computational cost and to easily compare with other QST algorithms, we are restricted to the above fixed tetrahedral POVM.

All numerical experiments are run on the laptop with a single Intel(R) Core(TM) i5 CPU @ 2.50GHz with 16GB RAM, and a single NVIDIA GeForce GTX 1070 GPU with 8GB RAM. The traditional MLE-based QST algorithms are based on the Matlab platform and uses the CPU to accelerate operations, while the NN-QST is based on the Pytorch platform and uses GPU to accelerate operations. The codes are available at github:QST-NNGMs-FNN [43].

4.2 Fast convergence of the unified state-mapping methods

We focus on the following ten mapping methods:

- The decomposition method Chol$_\Delta$ with the complex lower triangular matrix (4).
- The $\mathcal{F}[\cdot]$ and $\mathcal{S}[\cdot]$ projection methods (7).
- The simple unified method Chol$_H$ (14).
- The unified method based the $P$-order absolute projection $A[\cdot]_P$ (17), with $P = 0.5, 1, 1.5, 2, 3, 4$,

and test how the choice of state-mapping methods affects the convergence speed of the NN-QST.
As displayed in Fig. 2, the numerical results show our unified methods CholH and $P$-order absolute projection $A[^\cdot]_P$ with $P = 2$ admits the advantages in convergence speed and tomography accuracy, as both can converge to the perfect fidelity of 1 within 100 seconds on all tested states. Indeed, these two methods are mathematically identical, whereas it is easier to calculate the state fidelity with CholH, thus leading to a faster convergence. Although the decomposition method Chol$_\Delta$ has the fastest convergence speed in the early stage, it achieves a worse state fidelity of less than 1 in the final estimate process. During the training process within 100 seconds, the projection method $F[^\cdot]$ only reconstructs the states with exponentially decaying eigenvalues with fidelity 1, while $S[^\cdot]$ fails for all states.

It is interesting to point out that as $P$ increases, the performance of the $A[^\cdot]_P$ projections on average tends to become better before $P \leq 2$ and then worse after $P \geq 2$. Hence, the critical point is $P = 2$. However, it is also found that in Fig. 2(c) the $A[^\cdot]_P$ with $P = 1.5$ performs better on the W states and in Fig. 2(d) the unified projection with $P = 1$ and 1.5 on the GHZi states. Hence, the flexible $P$ in the unified state-mapping method (17) can have a better performance in reconstructing different states.

### 4.3 State purity robustness of the unified state-mapping methods

The state purity is an important property of the state, encoding the structural information of the state. Thus, we further analyze the tomography accuracy of different state-mapping methods when the states with a wide range of purity are considered.

The NN-QST with ten state-mapping methods are experimented on 10-qubit states generated with random purity. As shown in Fig. 3, our unified methods CholH and $A[^\cdot]_P$ with $P = 2$ have excellent robustness against state purity in the sense that both can achieve the perfect fi-
composition and states initialized by uniformly distributed eigenvalues initialized by uniformly distributed purity, and (b-d) 10 random mixed states of product, W, and GHZi each mapping method, experiments are implemented with (a) 50 random mixed states with exponential decay of eigenvalues, and the purity has a great effect on state fidelity, while other methods only work well for special states with narrow range of purity and are sensitive to the varying state purity. Moreover, for the CholΔ decomposition and $\mathcal{F}[\cdot]$ projection, their quantum state fidelity is positively correlated with purity, except for mixed states with exponentially decaying eigenvalues, and the purity has a great effect on the final fidelity of the projection $S[\cdot]$. It is also noted that for the $\mathcal{A}[\cdot]_P$ projection method, there is the negative correlation between the fidelity and purity at $P \leq 2$ and positive correlation at $P \geq 2$. Consequently, it becomes insensitive to the state purity when $P$ is around 2. This tendency further suggests that if the purity information is known, then we use $\mathcal{A}[\cdot]_P$ with smaller $P$ for low-purity states and larger $P$ for high-purity states. And, if purity is an unknown parameter, then $\mathcal{A}[\cdot]_2$ is a good choice. This is the reason that we choose CholH as the default state-mapping method for the NN-QST.

![Figure 3](image)

Figure 3: The tomography fidelity of the NN-QST with ten mapping methods versus state purity on 10-qubit states. The purity of the tested state $\rho$ is $\text{Tr}[\rho^2]$, with the range being from 0 to 1 (the purity of the pure state is 1). For each mapping method, experiments are implemented with (a) 50 random mixed states with exponential decay of eigenvalues initialized by uniformly distributed purity, and (b-d) 10 random mixed states of product, W, and GHZi states initialized by uniformly distributed $p$, respectively, of which the purity corresponds to $p$ is $(1 - 1/2^{10})p^2 + 1/2^{10}$. The number of iterations is fixed at 1000.

4.4 Fast convergence of the NN-QST on a large number of qubits

We then test how fast our neural network and other QST algorithms can tomography various states with fidelity 0.99. In particular, it is shown in Fig. 4 that the NN-QST outperforms over other two algorithms in the convergence time for the states of above 7 qubits. Even for the states with a small number of qubits, our method can accomplish the tomography task within 1 minute, close to those of other two algorithms. Note that the CG-APG and iMLE algorithms are only tested up to 9-qubit experiments due to the huge time cost for a larger number of qubits, while the NN-QST still can. Surprisingly, it is able to achieve the tomography fidelity 0.99 for 11-qubit mixed states within 100 seconds. Thus, the network-based QST is more scalable to deal with the multi-qubit system. Moreover, it is found that the NN-QST admits a better performance on all tested states, further demonstrating the to-
mography stability on a wide range of states.

4.5 Noise robustness of the NN-QST

All above experiments are implemented without noise, however, it is impractical to be free from noise induced by finite measurements and experimental environment. In this subsection, we consider the tomography performance of three QST algorithms under two types of noise: statistical noise and depolarizing noise. The statistical noise arises from the difficulty in obtaining ideal measurement probabilities from a finite number of measurements, and the depolarizing noise has a certain probability to change target states into the maximally mixed state.

4.5.1 Statistical noise robustness

We examine the effects of different sample sizes on the tomography performance of three QST algorithms on 8-qubit random mixed states. All measurement probabilities are calculated with one product-structured POVM and a certain number of samples from sampling a multinomial distribution of measurement probabilities.

First, it is natural to find in Fig. 5(a) that more samples improves the tomography fidelity of all QST algorithms. The NN-QST achieves a higher average fidelity than the CG-APG and iMLE algorithms when the same number of samples is consumed. And, given the same high fidelity, it can save one to two orders of sample resources. Then, Fig. 5(b) shows that the time consumption of the NN-QST is two orders of magnitude less than other two algorithms, under the same samples condition. These results indicate that
the NN-QST completes fidelity convergence on more states with increasing samples and a slow decrease in time, while other two algorithms almost remain unchanged in time, indicating that the convergence is completed in few states. Finally, it is demonstrated that the tomography fidelity is positively correlated with state purity, as illustrated in Fig. 5(c). Given a fixed state purity, the NN-QST achieves a higher fidelity than other algorithms under the same number of samples. And the NN-QST achieves a fidelity of 0.99 on states of wider purity as the number of samples increases, while other algorithms can not reach fidelity of 0.99 on all states.

4.5.2 Depolarizing noise robustness

The tomography performances of three algorithms on 8-qubit pure states under different depolarizing noise strength are then investigated. The measured probability is obtained by measuring the pure states after the depolarizing channel, with the sample sizes of $10^4$, $10^5$, and $10^6$. The presence of depolarizing noise reduces the tomography fidelity for all algorithms, as depicted in Fig. 6 where the black dashed line describes the ideal fidelity with respect to the noise strength $\lambda$. For the product and W states, the NN-QST can have a tomography fidelity, which is closer to the ideal than other algorithms with a $\lambda$. And all of them performs well on the GHZ$i$ state. It is noted that the increased number of samples makes the tomography fidelity of each algorithm closer to the ideal fidelity for each $\lambda$.

Furthermore, the NN-QST takes one or two orders of magnitude less time than other algorithms to achieve a fidelity close to the ideal fidelity (plus or minus 0.005) or finish the iteration under various $\lambda$, as shown in Fig. 7. Again, for the product and W states, the NN-QST achieves the near-ideal fidelity faster than other algorithms. On the GHZ$i$ state, the three algorithms are not close to the ideal fidelity except for complete noise strength ($\lambda=1$), and the time consumption of the NN-QST is lower to maintain the same degree of closeness. Thus, the effect of depolarizing noise on the NN-QST is weaker than on other QST algorithms.

5 Conclusion and Outlook

We have presented the NN-QST which utilizes the unified state-mapping technique (17), to achieve the ultrafast reconstruction of multi-qubit states. In comparison to various state-mapping strategies and fast QST methods, numerous numerical results shows that our method has a fast convergence speed to achieve a high tomography fidelity and a great robustness to the state purity and noise. It is demonstrated that we are able to accomplish the tomography task...
needs more samples to reach a given fidelity and to the NN-QST, reconstructing low-purity states of 11-qubit states within 2 minutes on a laptop. For each sample size, experiments are with (a-c) 50 pure states with a uniformly distributed $(X_n \approx n)$ represents that the number of samples used by the $X$ algorithm is $10^n$). For each sample size, experiments are with (a-c) 50 pure states with a uniformly distributed $\lambda$ of the product, $W$, and GHZ$_r$ states, respectively. The black dotted line is the ideal fidelity $1 - (1 - 1/2^n)^n$, if the QST algorithm perfectly reconstructs the mixed state disturbed by depolarizing noise. The number of iterations is fixed at 500.

Figure 7: The time to reach a fidelity close to the ideal fidelity or complete the iteration as a function of noise strength $\lambda$. We consider three sample sizes of $10^4$, $10^5$, and $10^6$ (X$_n$ represents that the number of samples used by the X algorithm is $10^n$). For each sample size, experiments are with (a-c) 50 pure states with a uniformly distributed $\lambda$ of the product, $W$, and GHZ$_r$ states, respectively. The number of iterations is fixed at 500.

of 11-qubit states within 2 minutes on a laptop. Moreover, it has been found that with respect to the NN-QST, reconstructing low-purity states needs more samples to reach a given fidelity and the depolarizing noise linearly reduces its tomography fidelity.

There are many interesting questions left for the future work. For example, more than one single POVM and/or adaptive measurements could be used to improve the reconstruction fidelity for a more wide range of multi-qubit states. And generative models could further be introduced to enhance the learnability of the neural network and then to improve the NN-QST. Finally, it would be interesting to test our method on real experimental setups.

Acknowledgments

We thank Dr. Changkang Hu for useful discussions. This research is supported by the National Natural Science Foundation of China (61903027, 72171172); the National Key R&D Program of China (2018YFE0105000, 2018YFB1305304); the Shanghai Municipal Science and Technology Major Project (2021SHZDXZ0100); the Shanghai Municipal Commission of Science and Technology (1951113210, 19511132101); the Shanghai Municipal Science and Technology Fundamental Project (21JC1405400); and Shanghai Gaofeng & Gaoyuan Project for University Academic Program Development (22-8).
References

[1] Howard Barnum and Alexander Wilce. “Local tomography and the jordan structure of quantum theory”. Found. Phys. 44, 192–212 (2014).

[2] M. Baur, A. Fedorov, L. Steffen, S. Filipp, M. P. da Silva, and A. Wallraff. “Benchmarking a quantum teleportation protocol in superconducting circuits using tomography and an entanglement witness”. Phys. Rev. Lett. 108, 040502 (2012).

[3] Diogo Cruz, Romain Fournier, Fabien Gremion, Alix Jeannerot, Kenichi Komagata, Tara Tosic, Jarla Thiesbrummel, Chun Lam Chan, Nicolas Macris, Marc-André Dupertuis, et al. “Efficient quantum algorithms for ghz and w states, and implementation on the ibm quantum computer”. Adv. Quantum Technol. 2, 1900015 (2019).

[4] Hsin-Yuan Huang, Richard Kueng, and John Preskill. “Predicting many properties of a quantum system from very few measurements”. Nat. Phys. 16, 1050–1057 (2020).

[5] K. Banaszek, G. M. D’Ariano, M. G. A. Paris, and M. F. Sacchi. “Maximum-likelihood estimation of the density matrix”. Phys. Rev. A 61, 010304 (1999).

[6] Jaroslav Řeháček, Zden ěk Hradil, E. Knill, and A. I. Lvovsky. “Diluted maximum-likelihood algorithm for quantum tomography”. Phys. Rev. A 75, 042108 (2007).

[7] G. B. Silva, S. Glancy, and H. M. Vasconcelos. “Investigating bias in maximum-likelihood quantum-state tomography”. Phys. Rev. A 95, 022107 (2017).

[8] Bo Qi, Zhibo Hou, Li Li, Daoyi Dong, Guoyong Xiang, and Guangcan Guo. “Quantum state tomography via linear regression estimation”. Sci Rep 3, 3496 (2013).

[9] Bo Qi, Zhibo Hou, Yuanlong Wang, Daoyi Dong, Han-Sen Zhong, Li Li, Guo-Yong Xiang, Howard M Wiseman, Chuan-Feng Li, and Guang-Can Guo. “Adaptive quantum state tomography via linear regression estimation: Theory and two-qubit experiment”. npj Quantum Inform. 3, 19 (2017).

[10] Yi-Lin Seah, Jiangwei Shang, Hui Khoon Ng, David John Nott, and Berthold-Georg Englert. “Monte carlo sampling from the quantum state space. ii”. New J. Phys. 17, 043018 (2015).

[11] Joseph M Lukens, Kody JH Law, Ajay Jasra, and Pavel Lougovski. “A practical and efficient approach for bayesian quantum state estimation”. New J. Phys. 22, 063038 (2020).

[12] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. “Quantum state tomography via compressed sensing”. Phys. Rev. Lett. 105, 150401 (2010).

[13] Carlos A Riofrio, David Gross, Steven T Flammia, Thomas Monz, Daniel Nigg, Rainer Blatt, and Jens Eisert. “Experimental quantum compressed sensing for a seven-qubit system”. Nat. Commun. 8, 15305 (2017).

[14] Anastasios Kyrillidis, Amir Kalev, Do hyung Park, Srinadh Bhojanapalli, Constantine Caramanis, and Sujay Sanghavi. “Provable compressed sensing quantum state tomography via non-convex methods”. npj Quantum Inform. 4, 36 (2018).

[15] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwenmer, and H. Weinfurter. “Permutationally invariant quantum tomography”. Phys. Rev. Lett. 105, 250403 (2010).

[16] Matthias Christandl and Renato Renner. “Reliable quantum state tomography”. Phys. Rev. Lett. 109, 120403 (2012).

[17] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. “Efficient quantum state tomography”. Nat. Commun. 1, 149 (2010).

[18] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio. “Scalable reconstruction of density matrices”. Phys. Rev. Lett. 111, 020401 (2013).

[19] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum. “Quantum state tomography with conditional generative adversarial networks”. Phys. Rev. Lett. 127, 140502 (2021).
[20] Hailan Ma, Daoyi Dong, Ian R. Petersen, Chang-Jiang Huang, and Guo-Yong Xiang. “A comparative study on how neural networks enhance quantum state tomography” (2021). arXiv:2111.09504.

[21] Dominik Koutný, Libor Motka, Zdeněk Hradil, Jaroslav Reháček, and Luís L. Sánchez-Soto. “Neural-network quantum state tomography”. Phys. Rev. A 106, 012409 (2022).

[22] John A. Smolin, Jay M. Gambetta, and Graeme Smith. “Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise”. Phys. Rev. Lett. 108, 070502 (2012).

[23] Eliot Bolduc, George C Knee, Erik M Gauger, and Jonathan Leach. “Projected gradient descent algorithms for quantum state tomography”. npj Quantum Inform. 3, 44 (2017).

[24] Vikesh Siddhu. “Maximum a posteriori probability estimates for quantum tomography”. Phys. Rev. A 99, 012342 (2019).

[25] Jiangwei Shang, Zhengyun Zhang, and Hui Khoon Ng. “Superfast maximum-likelihood reconstruction for quantum tomography”. Phys. Rev. A 95, 062336 (2017).

[26] Qian Xu and Shuqi Xu. “Neural network state estimation for full quantum state tomography” (2018). arXiv:1811.06654v2.

[27] Tao Xin, Sirui Lu, Ningping Cao, Galit Anikeeva, Dawei Lu, Jun Li, Guilu Long, and Bei Zeng. “Local-measurement-based quantum state tomography via neural networks”. npj Quantum Inform. 5, 109 (2019).

[28] Adriano Macarone Palmieri, Egor Kovlakov, Federico Bianchi, Dmitry Yudin, Stanislav Straupe, Jacob D Biamonte, and Sergey Kulessa. “Experimental neural network enhanced quantum tomography”. npj Quantum Inform. 6, 20 (2020).

[29] Giuseppe Carleo and Matthias Troyer. “Solving the quantum many-body problem with artificial neural networks”. Science 355, 602–606 (2017).

[30] Giacomo Torlai, Guglielmo Mazzaola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. “Neural-network quantum state tomography”. Nat. Phys. 14, 447–450 (2018).

[31] E. S. Tiunov, V. V. Tiunova (Vyborova), A. E. Ulanov, A. I. Lvovsky, and A. K. Fedorov. “Experimental quantum homodyne tomography via machine learning”. Optica 7, 448–454 (2020).

[32] Huikang Huang, Haozhen Situ, and Shenggen Zheng. “Bidirectional information flow quantum state tomography”. Chin. Phys. Lett. 38, 040303 (2021).

[33] Stewart Morawetz, Isaac J. S. De Vlugt, Juan Carrasquilla, and Roger G. Melko. “U(1)-symmetric recurrent neural networks for quantum state reconstruction”. Phys. Rev. A 104, 012401 (2021).

[34] Andrea Rocchetto, Edward Grant, Sergii Strelchuk, Giuseppe Carleo, and Simone Severini. “Learning hard quantum distributions with variational autoencoders”. npj Quantum Inform. 4, 28 (2018).

[35] Peter Cha, Paul Ginsparg, Felix Wu, Juan Carrasquilla, Peter L McMahon, and Eun-Ah Kim. “Attention-based quantum tomography”. Mach Learn: Sci. Technol. 3, 01LT01 (2021).

[36] A I Lvovsky. “Iterative maximum-likelihood reconstruction in quantum homodyne tomography”. J. Opt. B-Quantum Semicl. Opt. 6, S556–S559 (2004).

[37] Tillmann Baumgratz, Alexander Nüßeler, Marcus Cramer, and Martin B Plenio. “A scalable maximum likelihood method for quantum state tomography”. New J. Phys. 15, 125004 (2013).

[38] Doglas S. Gonçalves, Márcia A. Gomes-Ruggiero, and Carlile Lavor. “Global convergence of diluted iterations in maximum-likelihood quantum tomography”. Quantum Info. Comput. 14, 966–980 (2014).

[39] Ahmad Farooq, Uman Khalid, Junaid ur Rehman, and Hyundong Shin. “Robust quantum state tomography method for quantum sensing”. Sensors 22, 2669 (2022).

[40] Anna Keselman, Yimon Glickman, Nitzan Akerman, Shlomi Kotler, and Roee Ozeri.
A NN-QST Based on CNN

To explain the reason for choosing the FNN in Sec. 3.1, here, we replace the FNN with a CNN with nearly the same amount of parameters to perform the NN-QST, and other experimental setups such as the negative log-likelihood loss function, Rprop optimizer with 0.001 learning rate, simple unified state-mapping strategy Chol4, and one product-structured POVM, etc. are consistent with Sec. 4.4.

The CNN structure for the 6-qubit experiments used here is shown in Fig. 8, the CNN structure presented below is extended to $N$ qubits for full measurements. The number of full measurements is $K^N$, where $K$ is the number of elements of a single-qubit POVM, which is 4 here. To satisfy the convolution operation, we convert the input probability distributions to matrix form. The convolution layer has a kernel of size $3 \times 3$, stride length of 1, padding length of 1, 10 feature maps, and a rectified linear unit (ReLU) activation function. The max-pooling layer has a kernel of size $2N-3 \times 2N-3$. The fully connected layer has an input layer of 640 neurons, a hidden layer of $4N$ neurons, and an output of size $K^N$. The input and output of the CNN are consistent with the FNN in Sec. 3.1.

To illustrate the performance difference between the FNN and CNN, we compare the convergence time of the two mapping networks to achieve fidelity of 0.99 on various qubits. It is shown in Fig. 9 that the FNN outperforms over the CNN for the most states of above 8 qubits. Even for the states with a small number of qubits, the FNN can accomplish the tomography task within 1 minute, close to those of the CNN. Thus, the FNN is more suitable to deal with the large-qubit system.

Considering the tomographic requirements of large qubits and the loss of scalability by using the CNN to transform one-dimensional input data, we finally chose a 3-layer FNN as the mapping network.
Figure 9: The convergence time for the FNN (red) and CNN (orange) implementations to reach fidelity of 0.99 versus the number of qubits. For each network, experiments are implemented with (a) 50 random mixed states with exponential decay of eigenvalues initialized by uniformly distributed purity, and (b-d) 50 random mixed states of product, W, and GHZi states initialized by uniformly distributed $p$, respectively. A data point is the average convergence time over 50 states and the shaded area represents the standard deviation around the mean. The number of iterations is fixed at 1000.

And, the architecture of the CNN guarantees that a smaller number of parameters can be maintained under the same tomographic performance, which will be exploited in future research.