Quantum State Tomography via Nonconvex Riemannian Gradient Descent

Ming-Chien Hsu,[1,†] En-Jui Kuo,[1,2] Wei-Hsuan Yu,[3] Jian-Feng Cai,[4] and Min-Hsiu Hsieh[1,*]
[1]Hon Hai Quantum Computing Research Center, Taipei, Taiwan
[2]Joint Center for Quantum Information and Computer Science, NIST and University of Maryland, College Park, Maryland, USA
[3]Department of Mathematics, National Central University, Taoyuan, Taiwan
[4]Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

(Received 13 February 2023; revised 28 March 2024; accepted 15 April 2024; published 13 June 2024)

The recovery of an unknown density matrix of large size requires huge computational resources. State-of-the-art performance has recently been achieved with the factored gradient descent (FGD) algorithm and its variants since they are able to mitigate the dimensionality barrier by utilizing some of the underlying structures of the density matrix. Despite the theoretical guarantee of a linear convergence rate, convergence in practical scenarios is still slow because the contracting factor of the FGD algorithms depends on the condition number κ of the ground truth state. Consequently, the total number of iterations needed to achieve the estimation error ε can be as large as $O(\sqrt{\kappa}\ln(1/\varepsilon))$. In this Letter, we derive a quantum state tomography scheme that improves the dependence on κ to the logarithmic scale. Thus, our algorithm can achieve the approximation error ε in $O(\ln(1/\kappa\varepsilon))$ steps. The improvement comes from the application of nonconvex Riemannian gradient descent (RGD). The contracting factor in our approach is thus a universal constant that is independent of the given state. Our theoretical results of extremely fast convergence and nearly optimal error bounds are corroborated by the numerical results.

DOI: 10.1103/PhysRevLett.132.240804

The aim of quantum state tomography is to recover an unknown density matrix ρ of size $d \times d$ from the measurement outcome $y \in \mathbb{R}^m$. Since the dimension of the density matrix $d = 2^n$ grows exponentially with the qubit number n, the complexity of the reconstruction increases very quickly. The fact that, up to now, experimental demonstrations of tomography have only been performed for small qubit numbers [1,2], shows the difficulties.

In the context of large matrix dimensions d, the challenge arises from the substantial number of samples required to reconstruct a state, leading to the sample complexity problem of determining the necessary and sufficient copies of ρ. For trace distance tomography aiming for a trace distance error of δ, studies by O'Donnell and Wright [3,4] established that $O(rd/\delta^2)$ input samples suffice for reconstructing density matrices of rank r. Conversely, Wright [5] indicated that $\Omega(rd)$ samples are necessary when δ is held constant. On a different note, Yu [6] demonstrated that $O(10^n/\delta^2)$ copies are sufficient using Pauli measurements. In the case of fidelity tomography aiming for a description with $1-\epsilon$ fidelity, Haah *et al.* [7] showed that $O(dr\log(d/\epsilon)/\epsilon)$ copies of ρ are sufficient. This is complemented by the corresponding lower bound of $\Omega(rd/(\epsilon\log(d/r\epsilon)))$, which was later improved to $\Omega(rd/\epsilon)$ in the work of Yuen *et al.* [8].

However, the challenge extends beyond sample complexity. Time complexity, governed by the algorithm used for state estimation, significantly impacts tomography quality. Computational tasks involving the entire matrix become exceedingly slow with large system sizes. Finding more efficient algorithms is crucial for practical applications. Many standard and state-of-the-art algorithms require solving the eigensystems or the application of singular value decomposition (SVD), especially when they involve projection related to the eigenspectrum [9–11], singular value contracting operator [12–14], or a unitary transformation of eigenbasis [15,16]. Both SVD and eigenvalue decomposition possess time complexity $O(d^3)$ and thus can be slow. Some advanced approach can reduce the complexity but still has complexity $O(d^3)$ [17]. Other time-consuming operations involve the full $d \times d$ matrix, including Hessian calculation [18] and matrix inverse method [19–21]. Although computing a less costly proxy for the Hessian matrix can improve the efficiency [18], it is still a heuristic approach along with the initial costly rapid descent calculation and still provides no theoretical guarantee of performance and convergence.

More practical approaches include incorporating local information [22,23], adopting neural network architecture [22,24–29], or restricting the form of states [23,30,31]. However, these methods usually have no convergence guarantees or rely on some state representation assumptions. Another possible improvement in efficiency can be obtained by adopting the graphical-processing unit (GPU) for processing [32]. However, this relies on the hardware and the use of full linear matrix inversion so is not an

efficient approach from the algorithmic point of view. Without utilization of the structures behind the matrix, these algorithms tend to be slow when the system is large.

Time complexity can be reduced if the underlying structure of the matrices can be utilized. Since the density matrices of interest are mostly of low rank [33–38], nonconvex approaches encoding the rank structure inherent to the algorithm can perform much better [39–44]. In particular, the nonconvex projected factored gradient descent (FGD) approach [45], along with both the momentum-inspired version (MiFGD) [46] and the stochastic version [47], which maintain the low rank r structures by decomposing each density matrix as $\rho = \tilde{U}\tilde{U}^\dagger$ for $\tilde{U} \in \mathbb{C}^{d\times r}$, have been adopted for tomography. Indeed, faster estimation of quantum states has been achieved by the MiFGD variant [46], than the state-of-the-art convex [48–50] or nonconvex [51] algorithms, including recent deep learning approaches [52–55]. This process, however, ignores the eigenvalue dependence during factorization; therefore, each update is heavily dependent on the condition number κ of the underlying matrix. Moreover, the minimization of errors in each step is related to the eigenvalues and the contracting factor being close to 1. Therefore, it still takes numerous iterations to obtain the final estimation.

In this Letter, we use a much more efficient nonconvex Riemannian gradient descent (RGD) algorithm [56] that can overcome these difficulties, while still maintaining high guaranteed accuracy. The RGD algorithm has proven to be useful and efficient in both matrix recovery problems [62] and matrix completion problems [63]. Its success comes from suitably taking care of the eigenvalues (or singular values in general) in each iteration, so that much more efficient convergence can be expected, while still maintaining high accuracy. The results show that it takes logarithmic steps to achieve the desired accuracy and that nearly optimal error bounds under noise are guaranteed.

Problem formulation.—Consider a n-qubit system wherein the density matrix $\rho \in \mathbb{H}_d(\mathbb{C})$ is of dimension $d = 2^n$, where $\mathbb{H}_d(\mathbb{C}) = \{M | M \in \mathbb{C}^{d\times d}, M = M^\dagger\}$ is the Hermitian space. Every n-qubit density matrix can be expanded by using Pauli observables $\{W_i : i \in [d^2]\}$, each of which is in the form of $P_1 \otimes P_2 \otimes \cdots \otimes P_n$, where $\otimes$ means tensor product and each P_i is a 2 by 2 matrix chosen from the Pauli matrices $\{\mathbb{I}_{2\times 2}, \sigma_x, \sigma_y, \sigma_z\}$. In the Pauli basis expansion of ρ, each W_i has a corresponding coefficient $\mathrm{Tr}(W_i\rho)/d$.

For the quantum system of interest, we consider that the density matrix $\rho \in \mathbb{H}_d(\mathbb{C})$ has rank r. The input data used to reconstruct ρ are considered to be from the common Pauli measurement, that is, from the collection of $\mathrm{Tr}(W_i X)$ with the corresponding Pauli observables. Since the ρ of rank r only has $(2d - r)r$ degrees of freedom, only $m = O(rd) \ll d^2$ coefficients are needed to reconstruct the matrix, leading to the problem of compressed sensing in the tomography [34,35,64–68]. By choosing m basis elements $\{S_1, S_2, \ldots, S_m\}$ uniformly at random from the Pauli basis set $\{W_1, W_2, \ldots, W_{d^2}\}$, we define the linear (sensing) map $\mathcal{A}: \mathbb{H}_d(\mathbb{C}) \to \mathbb{R}^m$ with its ith component corresponding to S_i as

$$\big(\mathcal{A}(X)\big)_i = \sqrt{\frac{d}{m}}\mathrm{Tr}(S_i X), \tag{1}$$

for $X \in \mathbb{H}_d(\mathbb{C})$. The corresponding adjoint operator, $\mathcal{A}^\dagger: \mathbb{R}^m \to \mathbb{H}_d(\mathbb{C})$, has its action on $y \in \mathbb{R}^m$ as

$$\mathcal{A}^\dagger(y) = \sqrt{\frac{d}{m}}\sum_{i=1}^{m} y_i S_i. \tag{2}$$

The coefficient $\sqrt{d/m}$ is chosen to make the expected value $\mathbb{E}[\mathcal{A}^\dagger(\mathcal{A}(X))] = X$, a necessary condition for having the restricted isometric property (RIP) [69].

The linear map $\mathcal{A}$ obtained from Pauli measurement is guaranteed to have the RIP with a high probability [64], as long as $m = C \cdot rd\log^6 d$ for some constant $C = O(1/\delta_r^2)$, depending only on δ_r, where δ_r is the RIP constant for the matrices of rank r. Because of the guaranteed RIP of the Pauli measurement, matrices of rank r can be recovered through some suitable optimization approach.

Since measurement almost surely introduces errors, we write $y = \mathcal{A}(\rho) + z$ to denote the scaled coefficients carrying noise z. According to the concentration properties of the random variables [70–72], as long as the number of total measurements ml is large enough, the noise z can be bounded by $\|\mathcal{A}^\dagger(z)\| \le \lambda$, with a high probability for the desired bound λ, where $\|\cdot\|$ is the spectral norm [69].

Pauli operators for measurements $\{S_i\}_{i\in\{1,\ldots,m\}}$ are sampled only once before the measurements. Once $\{S_i\}$ are sampled, fixing operators $\mathcal{A}$ and $\mathcal{A}^\dagger$ simultaneously, they are utilized to obtain the local measurement results $y = (y_i)_{i=1\cdots m} \in \mathbb{R}^m$. Both $\{S_i\}$ and $y = (y_i)_{i=1\cdots m}$ serve as inputs for the algorithms and remain fixed during the execution of algorithms. We can then obtain an estimate $\hat{\rho}$ of the density matrix ρ from a given $y \in \mathbb{R}^m$, where each y_i corresponds to $(\mathcal{A}(X))_i$ with respect to S_i along with the noise z_i [69].

The tomography problem is formulated with y as the input, and relaxed to the following nonconvex optimization problem

$$\min_{X\in\mathbb{C}^{d\times d}} f(X) := \frac{1}{2}\|y - \mathcal{A}(X)\|_2^2 \quad \text{subject to rank } (X) \le r, \tag{3}$$

where the unit trace and semidefinite positive $X \succeq 0$ constraints are relaxed. The relaxation of the unit trace constraint is reasonable since the error distance between

Algorithm 1. RGD Algorithm.

Input: $\mathcal{A}$, y and rank r.
Initialize X_0 and do the singular value decomposition $X_0 = U_0\Sigma_0 V_0^\dagger$.
for $k = 1, \ldots$ **do**
1. find the direction $G_k = \mathcal{A}^\dagger(y - \mathcal{A}(X_k))$
2. determine the step size $\alpha_k = \|\mathcal{P}_{T_k}(G_k)\|_F^2 / \|\mathcal{A}(\mathcal{P}_{T_k}(G_k))\|_2^2$.
3. find an intermediate matrix $J_k = X_k + \alpha_k \mathcal{P}_{T_k}(G_k)$.
4. update the estimated matrix $X_{k+1} = \mathcal{H}_r(J_k)$
end for
Output: $\hat{\rho} = X_k$ when the stopping criteria is met.

the estimated $\hat{\rho}$ and the underlying ρ is small, along with the fact that the trace of $\hat{\rho}$ is influenced by the noise z and can deviate from 1. The condition $X \succeq 0$ is also relaxed since the eigenvalues and singular values are the same for the underlying to-be-solved ρ; therefore the final estimated $\hat{\rho}$ will automatically satisfy the positive semidefinite constraint. The noise bound λ is also not used as a constraint; the condition $\|\mathcal{A}^\dagger(z)\| \le \lambda$ is used to analyze the error bound of the final result.

Main result.—We first discuss the steps of the nonconvex RGD algorithm 1 applied to solve the tomography in detail, then demonstrate how the RGD can efficiently update the estimated solution with the error bound controlled as in Theorem 1. The number of required steps can be further deduced as shown in Corollary 1.

The RGD algorithm 1 is an iterative algorithm applied to solve the optimization problem (3) with a well-defined and fixed $\mathcal{A}$ which requires input y. It is basically a special type of projected gradient descent approach. In the kth iteration within the loop in Algorithm 1, the direction $G_k = -\nabla f(X_k)$ in step 1 follows the usual gradient. With the choice of a suitable step size α_k in step 2, the current estimated matrix X_k is updated along the gradient projected into the tangent space T_k to give W_k in step 3. Supposing that X_k has singular value decomposition (SVD) $X_k = U_k\Sigma_k V_k^\dagger$, then the projections onto its column and row spaces are denoted by $P_{U_k} = U_k U_k^\dagger$ and $P_{V_k} = V_k V_k^\dagger$ respectively. The tangent space T_k at the current step is determined by

$$T_k = \{X \in \mathbb{H}_d | (\mathbf{I} - P_{U_k}) X (\mathbf{I} - P_{V_k}) = 0\},$$

and the corresponding projection $\mathcal{P}_{T_k}$ is

$$\mathcal{P}_{T_k} : X \mapsto P_{U_k} X + X P_{V_k} - P_{U_k} X P_{V_k}.$$

The trick in step 4 is to apply the hard thresholding operator $\mathcal{H}_r$ to get the updated $(k+1)$th step matrix X_{k+1} which is still of rank r. No further normalization is needed. The operator $\mathcal{H}_r$ acts on any matrix X to produce a truncated rank r approximation X_r which preserves the top r singular values $\sigma_1, \sigma_2, \cdots \sigma_r$ and the corresponding singular vectors. All the remaining $\sigma_{r+1}, \cdots$ are discarded under $\mathcal{H}_r$.

A significant advantage of the RGD algorithm lies in the fact that the SVD in each iteration step only needs to be performed for matrices of size $2r \times 2r$, as opposed to $d \times d$, since J_k has rank at most $2r$ [62,69]. This efficiency is crucial because the time complexity of SVD for matrices of dimensions $\eta \times \eta$ is $O(\eta^3)$, and in most relevant cases, matrices have low rank $r \ll d = 2^n$.

In each iteration of the RGD algorithm 1, the step size α_k is determined from an exact line search, since the object function is quadratic over the set of matrices. In contrast, the FGD-type algorithms factorize each matrix X in the form of $\tilde{U}\tilde{U}^\dagger$ such that the objective function in the factored matrix $\tilde{U}$ is quartic. This makes it impossible to do an exact line search, meaning that some prior knowledge or parameters are required in FGD.

The power of the RGD to solve the tomography problem is demonstrated in Theorem 1 and Corollary 1. Let ρ be the ground truth density matrix of rank r with a measurement result $y = \mathcal{A}(\rho) + z \in \mathbb{R}^m$ where the mapping $\mathcal{A}$ is defined as in Eq. (1), and the noise z is supposed to obey $\|\mathcal{A}^\dagger(z)\| \le \lambda$. Denote the condition number of ρ to be $\kappa := \sigma_1/\sigma_r$, where σ_1 and σ_r denote the first and the rth singular values of ρ, respectively. Starting from the initial point $X_0 = \mathcal{H}_r(\mathcal{A}^\dagger(y))$ [73], the final estimated $\hat{\rho}$ can be arbitrarily close to the ground truth ρ for the noiseless case. We can provide an upper bound for the error of the estimated $\hat{\rho}$ in terms of λ, which goes back to the noiseless case for $\lambda = 0$ (or $z = 0$).

Theorem 1.—There exist constants $C_1, C_2 > 0$ such that, when provided with $\lambda \le C_1\sigma_r/\sqrt{r}$ and $m \ge C_2\kappa^2 r^2 d \log^6 d$, then the kth iteration of the RGD algorithm 1 with an initial point $X_0 = \mathcal{H}_r(\mathcal{A}^\dagger(y))$ has a rank of at most r and is guaranteed to be close to the true ρ in the Frobenius norm distance bounded by

$$\|X_k - \rho\|_F \le \|X_0 - \rho\|_F \cdot \bar{\gamma}^k + \frac{2\sqrt{2r}\lambda}{1 - \delta_{3r}} \left(\frac{1}{1 - \bar{\gamma}}\right), \quad (4)$$

where the contracting factor $\bar{\gamma} < 1$ is a universal bound in all steps and δ_{3r} is the RIP constant of $\mathcal{A}$.

Proof.—The outline of the proof starts by finding a recursion relation between X_{k+1} and X_k. First, according to the triangular inequality, we have

$$\begin{aligned}\|X_{k+1} - \rho\|_F &\le \|X_{k+1} - W_k\|_F + \|W_k - \rho\|_F \\ &\le 2\|\rho - W_k\|_F,\end{aligned}$$

where $\|X_{k+1} - W_k\|_F \le \|W_k - \rho\|_F$ comes from the Eckart-Young theorem. Since W_k is in the tangent space T_k of X_k, this is further related to a B_k term in terms of $\|X_k - \rho\|_F$ plus another noise term B_z in terms of λ. The recursion relation makes the noise propagate. The sufficient

conditions can then be quantified so that each update has a contracting factor of $\bar{\gamma} < 1$ in the B_k term and the accumulated B_z term can be upper bounded [69]. ∎

The error is reduced by the application of a multiplicative contracting factor $\bar{\gamma}$ in each update, leading to a favorable linear convergence rate. Specifically, the factor $\bar{\gamma}$ in our RGD algorithm is a universal constant, which is independent of all parameters, including the RIP constant, the condition number κ of the ground truth ρ and so on. In other words, the error decays exponentially with a constant factor.

Corollary 1.—Under the conditions in Theorem 1, there exist positive constants C_0, C_1, C_2 all being $O(1)$ and $C_1 < C_2$ such that the RGD algorithm 1 can output the estimated density matrix $\hat{\rho}$ close to ρ of rank r, obeying

$$\frac{\|\hat{\rho}-\rho\|_F}{\|\rho\|_F} \le C_2 \frac{\sqrt{r}\lambda}{\|\rho\|_F},$$

after

$$\frac{1}{\ln(1/\bar{\gamma})}\left(\ln\left(\frac{2C_0\|\rho\|_F}{r\kappa\lambda}+2\sqrt{2}\right)-\ln(C_2-C_1)\right) \quad (5)$$

iteration steps, where $\kappa := \sigma_1/\sigma_r$ is the condition number of ρ, and $\bar{\gamma}$ is a universal constant smaller than 1.

When applied to the noiseless case, that is $\lambda = 0$, the RGD algorithm outputs $\hat{\rho}$ with

$$\frac{\|\hat{\rho}-\rho\|_F}{\|\rho\|_F} \le \varepsilon,$$

after $\ln(C_0/\sqrt{r}\kappa\varepsilon)/\ln(1/\bar{\gamma})$ iteration steps.

The corollary demonstrates two advantages. In terms of the time, the convergence is extremely fast, with the number of steps required to achieve the final error ε being on the order of $O(\ln(1/\kappa\varepsilon))$. This logarithmic dependence on κ in the convergence steps is an exponential improvement over the FGD algorithm and its variants. Although the FGD-type algorithms can also achieve a linear convergence rate, their iterative contracting factor is not universal. Its form can be written as $1-(1/\kappa^{\alpha})$, where α can be improved to 0.5 in some variants. This gives a total number of iteration steps of $O(\kappa^{\alpha}\ln(1/\varepsilon))$ to achieve the final error ε.

From the aspect of the estimation error, the RGD estimated matrix $\hat{\rho}$ is close to the ground truth ρ with a nearly optimal error distance, under the following conditions: small noise $\lambda \le C_1\sigma_r/\sqrt{r}$ and a large enough sample $m \ge C_2\kappa^2 r^2 d\log^6 d$. The final achievable error bound depends on the initial input noise $z \in \mathbb{R}^m$. In the noiseless case, where $z = 0$, the error can be reduced to nearly zero with arbitrary precision. In the noisy case, the RGD result shows the error bound to be in the Frobenius norm $\|\hat{\rho}-\rho\|_F \le C\sqrt{r}\lambda$, which is tighter than the more commonly seen nuclear norm. By virtue of triangular inequality, this bound can be converted to the nuclear norm $\|\hat{\rho}-\rho\|_* \le Cr\lambda$. Both are of the same order as the best theoretical results obtained from the convex optimization approaches [71,76], and hence are nearly optimal.

Numerical experiment.—As FGD-type algorithms have demonstrated superior performance compared to existing state-of-the-art methods, to the best of our knowledge, we conducted a numerical comparison between our RGD approach and the MiFGD algorithm [46], an improved variant of the original FGD [45]. To ensure a fair comparison, both RGD and MiFGD were executed to recover the same ground truth density matrices ρ [77].

Three types of density matrices are demonstrated: (a) the GHZ state [78]: $\mathrm{GHZ}(n) = [(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})/\sqrt{2}]$, (b) the Hadamard state: $\mathrm{Hadamard}(n) = [(|0\rangle + |1\rangle)/\sqrt{2}]^{\otimes n}$, and (c) the randomly generated mixed states. For pure states (a) and (b), the quantum states ρ are generated by using the open-source software Qiskit [79], with the vectors y obtained from the frequency counting result of l shots measurements [69]. For mixed states (c), random density matrices are generated by using the QuTip package [80,81], along with the exact value of the vector y being used for matrix recovery.

The numerical results regarding the scaling of computation time with respect to the qubit number n are

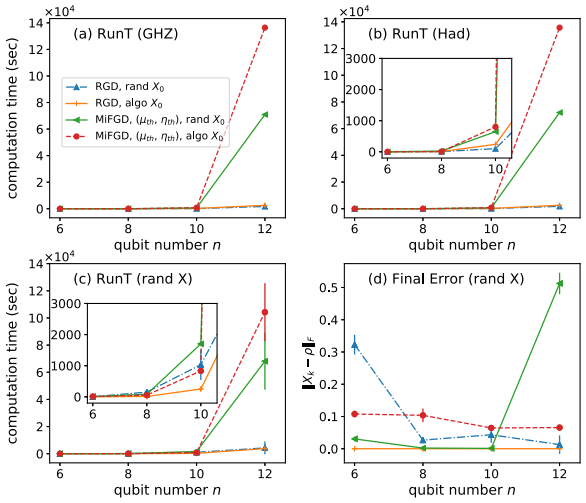

FIG. 1. (a),(b), and (c) The computation time scaling with respect to the qubit number n for the (a) GHZ state, (b) Hadamard state, and (c) random mixed states of rank 3, respectively. (d) The final target error $\|X_k-\rho\|_F$ for random mixed states corresponding to (c). Both (c) and (d) are plotted using the average value with standard deviation over 5 independent samples. The label "algo X_0" refers to each algorithm designed initial X_0, while the "rand X_0" is the random initial X_0 chosen the same for both RGD and MiFGD algorithms. Here the theoretical value of parameters (μ,η) is used for demonstration in the MiFGD, while RGD is parameter free. In (d), it is evident that the RGD algorithm algo X_0 initialization consistently achieves the lowest final target error $\|X_k-\rho\|_F$, reaching as low as $10^{-4}\sim 10^{-5}$ in most cases.

summarized in Fig. 1. The algorithms converge based on the iteration relative error, calculated as $\|X_{k+1} - X_k\|_F / \|X_k\|$ or $\|U_{k+1} - U_k\|_F / \|U_k\|$, being less than the stopping criteria (10^{-4} or 10^{-3}) during the iterations. The criterion is applied instead of a target error, as the algorithms assume the target density matrix ρ to be unknown. It is apparent that the RGD with the algorithm initialization $X_0 = \mathcal{H}_r(\mathcal{A}^\dagger(y))$ requires the shortest convergence time to achieve the smallest final target error. Consequently, the numerical results corroborate the superiority of the RGD algorithm in terms of both convergence time and final target error. Further simulation results, including different choices of parameters (μ, η) for the MiFGD and testing various condition number κ cases, support this conclusion [69].

Summary.—We show that the tomography problem can be efficiently solved using a Riemannian gradient descent (RGD) algorithm, because the RGD searches for the solution over a tangent space of low rank. The estimated matrix is updated iteratively with the error minimized by a multiplicative contracting factor $\bar{\gamma}$ in each step. Specifically, the $\bar{\gamma}$ is universal and independent of the condition number κ, the rank, and so on. Therefore, the required steps are $O[\ln(1/\kappa\varepsilon)]$ to achieve the final error ε. In the noiseless case, the error can be arbitrarily small. For the noisy case, the regimes are quantified to theoretically prove that the RGD approach can converge to the ground truth density matrix with a nearly optimal bound difference. The numerical results corroborate that the RGD outperforms the current state-of-the-art FGD-type approaches.

The authors are grateful to the referees for their valuable suggestions, which have helped enhance the comprehensiveness and conclusiveness of the work. W.-H. Y. is supported by National Science and Technology Council (NSTC) under Grant No. 109-2628-M-008-002-MY4. J.-F. C. is supported by Hong Kong Research Grant Council (HKRGC) General Research Fund (GRF) 16310620, 16306821, and 16307023.

*Corresponding author: min-hsiu.hsieh@foxconn.com
†Corresponding author: hubery.mc.hsu@foxconn.com

[1] Christian Schwemmer, Géza Tóth, Alexander Niggebaum, Tobias Moroder, David Gross, Otfried Gühne, and Harald Weinfurter, Experimental comparison of efficient tomography schemes for a six-qubit state, Phys. Rev. Lett. **113**, 040503 (2014).
[2] 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**, 1 (2017).
[3] Ryan O'Donnell and John Wright, Efficient quantum tomography, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 2016), pp. 899–912, 10.1145/2897518.2897544.
[4] Ryan O'Donnell and John Wright, Efficient quantum tomography II, in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 2017), pp. 962–974, 10.1145/3055399.3055454.
[5] John Wright, How to learn a quantum state, Ph.D. thesis, Carnegie Mellon University, 2016.
[6] Nengkun Yu, Sample efficient tomography via pauli measurements, arXiv:2009.04610.
[7] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu, Sample-optimal tomography of quantum states, in *Proceedings of the 48th annual ACM symposium on Theory of Computing* (2017), pp. 913–925, 10.1145/2897518.2897585.
[8] Henry Yuen, An improved sample complexity lower bound for (fidelity) quantum state tomography, Quantum **7**, 890 (2023).
[9] Jaroslav Řeháček, Zden ěk Hradil, E. Knill, and A. I. Lvovsky, Diluted maximum-likelihood algorithm for quantum tomography, Phys. Rev. A **75**, 042108 (2007).
[10] Douglas Soares Gonçalves, Márcia A Gomes-Ruggiero, and Carlile Lavor, A projected gradient method for optimization over density matrices, Optim. Methods Software **31**, 328 (2016).
[11] Eliot Bolduc, George C. Knee, Erik M. Gauger, and Jonathan Leach, Projected gradient descent algorithms for quantum state tomography, npj Quantum Inf. **3**, 1 (2017).
[12] Kai Zheng, Kezhi Li, and Shuang Cong, A reconstruction algorithm for compressive quantum tomography using various measurement sets, Sci. Rep. **6**, 1 (2016).
[13] Jiaojiao Zhang, Kezhi Li, Shuang Cong, and Haitao Wang, Efficient reconstruction of density matrices for high dimensional quantum state tomography, Signal Process. **139**, 136 (2017).
[14] Zhilin Hu, Kezhi Li, Shuang Cong, and Yaru Tang, Reconstructing pure 14-qubit quantum states in three hours using compressive sensing, IFAC-PapersOnLine **52**, 188 (2019).
[15] J. Řeháček, Z. Hradil, and M. Ježek, Iterative algorithm for reconstruction of entangled states, Phys. Rev. A **63**, 040303(R) (2001).
[16] 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).
[17] Chung-En Tsai, Hao-Chung Cheng, and Yen-Huan Li, Fast minimization of expected logarithmic loss via stochastic dual averaging, arXiv:2311.02557.
[18] Jiangwei Shang, Zhengyun Zhang, and Hui Khoon Ng, Superfast maximum-likelihood reconstruction for quantum tomography, Phys. Rev. A **95**, 062336 (2017).
[19] Bo Qi, Zhibo Hou, Li Li, Daoyi Dong, Guoyong Xiang, and Guangcan Guo, Quantum state tomography via linear regression estimation, Sci. Rep. **3**, 1 (2013).
[20] Kezhi Li and Shuang Cong, A robust compressive quantum state tomography algorithm using ADMM, IFAC Proc. Volumes **47**, 6878 (2014).

[21] Kezhi Li, Hui Zhang, Sen Kuang, Fangfang Meng, and Shuang Cong, An improved robust admm algorithm for quantum state tomography, Quantum Inf. Process. **15**, 2343 (2016).
[22] 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 Inf. **5**, 109 (2019).
[23] Yuchen Guo and Shuo Yang, Scalable quantum state tomography with locally purified density operators and local measurements, arXiv:2307.16381.
[24] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).
[25] Qian Xu and Shuqi Xu, Neural network state estimation for full quantum state tomography, arXiv:1811.06654.
[26] 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).
[27] Tobias Schmale, Moritz Reh, and Martin Gärttner, Efficient quantum state tomography with convolutional neural networks, npj Quantum Inf. **8**, 115 (2022).
[28] Dominik Koutný, Libor Motka, Zdeněk Hradil, Jaroslav Řeháček, and Luis L. Sánchez-Soto, Neural-network quantum state tomography, Phys. Rev. A **106**, 012409 (2022).
[29] Hailan Ma, Zhenhong Sun, Daoyi Dong, and Dong Gong, Learning informative latent representation for quantum state tomography, arXiv:2310.00518.
[30] T. Baumgratz, A. Nüßeler, M. Cramer, and M. B. Plenio, A scalable maximum likelihood method for quantum state tomography, New J. Phys. **15**, 125004 (2013).
[31] B. P. Lanyon, C. Maier, Milan Holzäpfel, Tillmann Baumgratz, C. Hempel, P. Jurcevic, Ish Dhand, A. S. Buyskikh, A. J. Daley, Marcus Cramer *et al.*, Efficient tomography of a quantum many-body system, Nat. Phys. **13**, 1158 (2017).
[32] Zhibo Hou, Han-Sen Zhong, Ye Tian, Daoyi Dong, Bo Qi, Li Li, Yuanlong Wang, Franco Nori, Guo-Yong Xiang, Chuan-Feng Li *et al.*, Full reconstruction of a 14-qubit state within four hours, New J. Phys. **18**, 083036 (2016).
[33] Steven T. Flammia, Andrew Silberfarb, and Carlton M. Caves, Minimal informationally complete measurements for pure states, Found. Phys. **35**, 1985 (2005).
[34] 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).
[35] David Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inf. Theory **57**, 1548 (2011).
[36] Charles H. Baldwin, Ivan H. Deutsch, and Amir Kalev, Strictly-complete measurements for bounded-rank quantum-state tomography, Phys. Rev. A **93**, 052105 (2016).
[37] Nan Li, Christopher Ferrie, Jonathan A. Gross, Amir Kalev, and Carlton M. Caves, Fisher-symmetric informationally complete measurements for pure states, Phys. Rev. Lett. **116**, 180402 (2016).
[38] Yu Wang and Yun Shang, Pure state 'really'informationally complete with rank-1 POVM, Quantum Inf. Process. **17**, 1 (2018).
[39] Tuo Zhao, Zhaoran Wang, and Han Liu, A nonconvex optimization framework for low rank matrix estimation, Adv. Neural Inf. Process. Syst. **28**, 559 (2015), https://proceedings.neurips.cc/paper/2015/hash/39461a19e9eddfb385ea76b26521ea48-Abstract.html.
[40] Srinadh Bhojanapalli, Anastasios Kyrillidis, and Sujay Sanghavi, Dropping convexity for faster semi-definite optimization, in *Proceedings of the 29th Annual Conference on Learning Theory*, edited by Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, Volume 49 of Proceedings of Machine Learning Research (Columbia University, New York, New York, USA, 2016), pp. 530–582.
[41] Ruoyu Sun and Zhi-Quan Luo, Guaranteed matrix completion via non-convex factorization, IEEE Trans. Inf. Theory **62**, 6535 (2016).
[42] Stephen Tu, Ross Boczar, Max Simchowitz, Mahdi Soltanolkotabi, and Ben Recht, Low-rank solutions of linear matrix equations via procrustes flow, in *Proceedings of The 33rd International Conference on Machine Learning, volume 48 of Proceedings of Machine Learning Research, New York, NY*, edited by Maria Florina Balcan and Kilian Q. Weinberger (PMLR, 2016), pp. 964–973, https://proceedings.mlr.press/v48/tu16.html.
[43] Rong Ge, Chi Jin, and Yi Zheng, No spurious local minima in nonconvex low rank problems: A unified geometric analysis, in *International Conference on Machine Learning, Sydney, Australia* (PMLR, 2017), pp. 1233–1242, https://proceedings.mlr.press/v70/ge17a.html.
[44] Dohyung Park, Anastasios Kyrillidis, Constantine Caramanis, and Sujay Sanghavi, Finding low-rank solutions via nonconvex matrix factorization, efficiently and provably, SIAM J. Imaging Sci. **11**, 2165 (2018).
[45] Anastasios Kyrillidis, Amir Kalev, Dohyung Park, Srinadh Bhojanapalli, Constantine Caramanis, and Sujay Sanghavi, Provable compressed sensing quantum state tomography via non-convex methods, npj Quantum Inf. **4**, 1 (2018).
[46] Junhyung Lyle Kim, George Kollias, Amir Kalev, Ken X. Wei, and Anastasios Kyrillidis, Fast quantum state reconstruction via accelerated non-convex programming, Photonics **10**, 116 (2023).
[47] Junhyung Lyle Kim, Mohammad Taha Toghani, César A Uribe, and Anastasios Kyrillidis, Local stochastic factored gradient descent for distributed quantum state tomography, arXiv:2203.11579.
[48] Alp Yurtsever, Quoc Tran Dinh, and Volkan Cevher, A universal primal-dual convex optimization framework, Adv. Neural Inf. Process. Syst. **28**, 3150 (2015).
[49] Steven Diamond and Stephen Boyd, Cvxpy: A Python-embedded modeling language for convex optimization, J. Mach. Learn. Res. **17**, 2909 (2016).
[50] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd, A rewriting system for convex optimization problems, J. Control Decis. **5**, 42 (2018).
[51] Elad Hazan, Sparse approximate solutions to semidefinite programs, in *Latin American Symposium on Theoretical Informatics* (Springer, New York, 2008), pp. 306–316.

[52] Xun Gao and Lu-Ming Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 1 (2017).
[53] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).
[54] Matthew J. S. Beach, Isaac De Vlugt, Anna Golubeva, Patrick Huembeli, Bohdan Kulchytskyy, Xiuzhe Luo, Roger G. Melko, Ejaaz Merali, and Giacomo Torlai, QuCumber: Wavefunction reconstruction with neural networks, SciPost Phys. **7**, 009 (2019).
[55] Giacomo Torlai and Roger G Melko, Machine-learning quantum states in the NISQ era, Annu. Rev. Condens. Matter Phys. **11**, 325 (2020).
[56] Some comparison with earlier approaches can be found in the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.240804, which includes Refs. [57–61] for showing the developement and relations of methods.
[57] 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).
[58] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956 (2010).
[59] Donald Goldfarb and Shiqian Ma, Convergence of fixed-point continuation algorithms for matrix rank minimization, Found. Comput. Math. **11**, 183 (2011).
[60] Prateek Jain, Raghu Meka, and Inderjit Dhillon, Guaranteed rank minimization via singular value projection, Adv. Neural Inf. Process. Syst. 23 (2010), https://papers.nips.cc/paper/2010/hash/08d98638c6fcd194a4b1e6992063e944-Abstract.html.
[61] Jared Tanner and Ke Wei, Normalized iterative hard thresholding for matrix completion, SIAM J. Sci. Comput. **35**, S104 (2013).
[62] Ke Wei, Jian-Feng Cai, Tony F. Chan, and Shingyu Leung, Guarantees of Riemannian optimization for low rank matrix recovery, SIAM J. Matrix Anal. Appl. **37**, 1198 (2016).
[63] Ke Wei, Jian-Feng Cai, Tony F. Chan, and Shingyu Leung, Guarantees of Riemannian optimization for low rank matrix completion, Inv. Prob. Imag. **14**, 233 (2020).
[64] Yi-Kai Liu, Universal low-rank matrix recovery from Pauli measurements, Adv. Neural Inf. Process. Syst. **24**, 1638 (2011), https://proceedings.neurips.cc/paper/2011/hash/e820a45f1dfc7b95282d10b6087e11c0-Abstract.html.
[65] Wei-Tao Liu, Ting Zhang, Ji-Ying Liu, Ping-Xing Chen, and Jian-Min Yuan, Experimental quantum state tomography via compressed sampling, Phys. Rev. Lett. **108**, 170403 (2012).
[66] A. Smith, C. A. Riofrío, B. E. Anderson, H. Sosa-Martinez, I. H. Deutsch, and P. S. Jessen, Quantum state tomography by continuous measurement and compressed sensing, Phys. Rev. A **87**, 030102(R) (2013).
[67] Amir Kalev, Robert L. Kosut, and Ivan H. Deutsch, Quantum tomography protocols with positivity are compressed sensing protocols, npj Quantum Inf. **1**, 1 (2015).
[68] Adrian Steffens, C. A. Riofrío, Will McCutcheon, Ingo Roth, Bryn A. Bell, Alex McMillan, M. S. Tame, J. G. Rarity, and Jens Eisert, Experimentally exploring compressed sensing quantum tomography, Quantum Sci. Technol. **2**, 025005 (2017).
[69] Additional discussions, such as the coefficient choice of $\mathcal{A}$, details of the proof, shot measurements, economic SVD in RGD, performance under varying condition numbers, and futher numerical results, are available in the Supplemental Material.
[70] Joel A. Tropp, User-friendly tail bounds for sums of random matrices, Found. Comput. Math. **12**, 389 (2012).
[71] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).
[72] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A. Tropp, Fast state tomography with optimal error bounds, J. Phys. A **53**, 204001 (2020).
[73] The hard thresholding operator $\mathcal{H}_r(\cdot)$ can be employed by the randomized SVD with details in the Supplemental Material and Refs. [74,75].
[74] Xu Feng, Wenjian Yu, and Yaohang Li, Faster matrix completion using randomized svd, in *Proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)* (IEEE, New York, 2018), pp. 608–615, https://ieeexplore.ieee.org/document/8576096.
[75] N. Halko, P. G. Martinsson, and J. A. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev. **53**, 217 (2011).
[76] Emmanuel J. Candes and Yaniv Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, IEEE Trans. Inf. Theory **57**, 2342 (2011).
[77] The codes of the RGD and MiFGD are built on the same basis. The RGD code is available at the github: https://github.com/Hon-Hai-Quantum-Computing/tomography_RGD.
[78] Scott N. Walck and David W. Lyons, Only n-qubit greenberger-horne-zeilinger states are undetermined by their reduced density matrices, Phys. Rev. Lett. **100**, 050501 (2008).
[79] QISKit Development Team, Qiskit, Qiskit: An open-source framework for quantum computing, 2021.
[80] J. Robert Johansson, Paul D. Nation, and Franco Nori, Qutip: An open-source python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **183**, 1760 (2012).
[81] J. R. Johansson, P. D. Nation, and Franco Nori, Qutip 2: A Python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **184**, 1234 (2013).