Quantum state tomography via non-convex Riemannian gradient descent

Ming-Chien Hsu[1], En-Jui Kuo[1,2], Wei-Hsuan Yu[3], Jian-Feng Cai[4], 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, Taiwan

[4]Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

Abstract

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 $O(\sqrt{\kappa}\ln(\frac{1}{\varepsilon}))$. In this work, 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 $O(\ln(\frac{1}{\kappa\varepsilon}))$ steps. The improvement comes from the application of non-convex 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.

Preliminary

- For quantum systems, the density matrix ρ is mostly of low rank. Consider ρ be of rank $r \ll d$. Then ρ only has $(2d-r)r$ d.o.f. We choose m basis elements $\{S_1, S_2, \cdots, S_m\}$ i.i.d. uniformly at random from the Pauli basis set $\{W_1, W_2, \cdots, W_{d^2}\}$.
- Define a linear (sensing) map $\mathcal{A} : \mathbb{H}_d(\mathbb{C}) \to \mathbb{R}^m$ with its i-th component $(\mathcal{A}(X))_i = \sqrt{\frac{d}{m}}\operatorname{Tr}(S_i X)$, The corresponding adjoint operator $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{H}_d(\mathbb{C})$ is $\mathcal{A}^*(y) = \sqrt{\frac{d}{m}}\sum_{i=1}^m y_i S_i$. Therefore, $\mathcal{A}^*\mathcal{A}(X) = \frac{d}{m}\sum_{i=1}^m \operatorname{Tr}(S_i X)S_i$. The expectation is then $\mathbb{E}[\mathcal{A}^*\mathcal{A}(X)] = \frac{1}{d}\sum_{i=1}^{d^2}\operatorname{Tr}(W_i X)W_i = X$.

Main Result

The density matrix ρ is considered of rank r with singular values $\{\sigma_i\}_{i=1}^r$ in decreasing order. Denote condition number of ρ be $\kappa := \sigma_1/\sigma_r$. The measured data is $y = \mathcal{A}(\rho) + z \in \mathbb{R}^m$, where the noise z is supposed to obey $\|\mathcal{A}^*(z)\| \le \lambda$.

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

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

where $\bar{\gamma} < 1$ is a universal contracting bound and δ_{3r} is the restricted isometry constant of $\mathcal{A}$.

Corollary 1. [2] There exist positive constants C_0, C_1, C_2 all being $O(1)$ and $C_1 < C_2$ such that the RGD algorithm 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 $\frac{1}{\ln(1/\bar{\gamma})}\left(\ln\left(\frac{2C_0\|\rho\|_F}{r\kappa\lambda} + 2\sqrt{2}\right) - \ln(C_2 - C_1)\right)$ iteration steps, where $\bar{\gamma}$ 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 $\ln\left(\frac{C_0}{\sqrt{r}\kappa\varepsilon}\right)/\ln(\frac{1}{\bar{\gamma}})$ iteration steps.

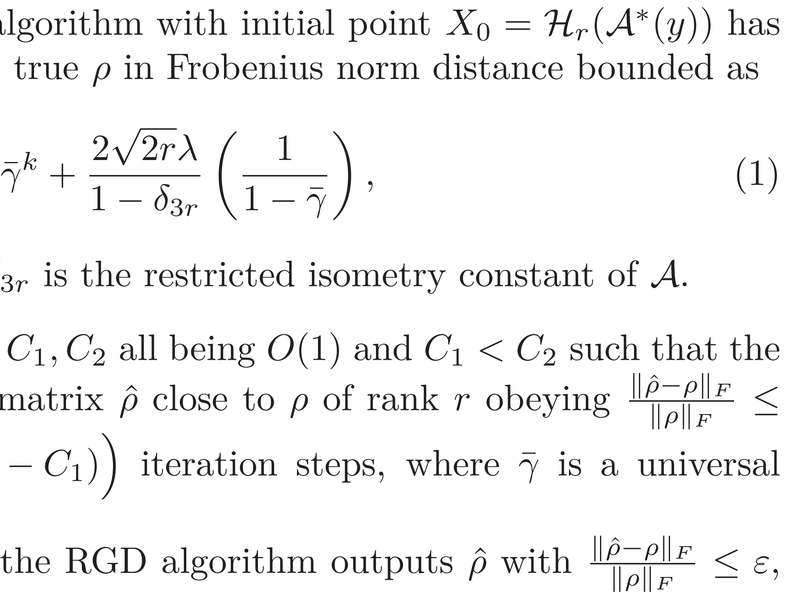
Numerical Setting and Result

- We use the open-source software Qiskit and IBM quantum simulator.
- We compare with the Momentum-Inspired Factored Gradient Decent (MiFGD):

$$\min_{A\in\mathbb{C}^{d\times r}} f(X) := \frac{1}{2}\|y - \mathcal{A}(AA^*)\|_2^2$$

  subject to $\|A\|_F^2 \le 1$, where $A \in \mathbb{C}^{d\times r}$. The hyperparameter momentum is $\mu \in \{1/8, 1/4, 1/3, 1/2, 3/4\}$ and the step size is $\eta = 0.01$ as [3].
- For both the Hadamard(6) state and the GHZ(6) state, we use $m = 819 \approx 0.2 \times 4^6$ Pauli measurements and $l = 2400$ shots. For both the Hadamard(8) state and the GHZ(8) state, we use $m = 3276 \approx 0.05 \times 4^8$ Pauli measurements and $l = 8600$ shots.
- The four demonstrated RGD cases all have convergence rates much faster than the MIFGD method for all μ.

Simulation Plot



The RGD algorithm is performed on the measured data (y_M) and the exact data (y_E) separately. The initial guess is either random X_0 or $X_0 = \mathcal{H}_r(\mathcal{A}^\dagger(y))$. The compared MiFGD algorithm is plotted in dots.

RGD Algorithm

The density matrix $\rho \in \mathbb{C}^{d\times d}$ is $\min_{X\in\mathbb{C}^{d\times d}} f(X) := \frac{1}{2}\|y - \mathcal{A}(X)\|_2^2$ subject to $\operatorname{rank}(X) \le r$

Algorithm 1 [1] K. Wei and J.-F. Cai and T. F Chan and S. Leung (2016)

Input: $\mathcal{A}, y$ and rank r.
Initialize X_0 and do the SVD $X_0 = U_0\Sigma_0 V_0^*$.
for $k = 1, \ldots$ **do**
1. find the direction $G_k = \mathcal{A}^*(y - \mathcal{A}(X_k))$
2. choose the step size $\alpha_k = \frac{\|\mathcal{P}_{T_k}(G_k)\|_F^2}{\|\mathcal{A}(\mathcal{P}_{T_k}(G_k))\|_2^2}$.
3. find a matrix $W_k = X_k + \alpha_k\mathcal{P}_{T_k}(G_k)$.
4. update the matrix $X_{k+1} = \mathcal{H}_r(W_k)$

end for
Output: $\hat{\rho} = X_k$ when the stopping criteria is met.

The tangent space T_k at the k-th step X_k is determined by

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

The corresponding projection is $\mathcal{P}_{T_k} : X \mapsto P_{U_k}X + XP_{V_k} - P_{U_k}XP_{V_k}$.

Conclusion

- The RGD approach is efficient in solving the quantum state tomography since it searches for possible solutions over low rank tangent space.
- In each iteration, the error is minimized by a universal contracting factor $\bar{\gamma} < 1$.
- The RGD approach can achieve the final result with nearly optimal error bound in Frobenius norm. In the noiseless case, the error can be arbitrarily small.
- Numerical simulations corroborate our results, and demonstrate that our approach performs better than state-of-the-art non-convex MIFGD approach.

References

[1] K. Wei, J.-F. Cai, T. F Chan, and S. Leung. *SIAM J. on Mat. Anal. and App.*, 37(3):1198–1222, 2016.

[2] M.-C. Hsu, E.-J/ Kuo, W.-H. Yu, J.-F. Cai, and M.-H. Hsieh. *arxiv:2210.04717.*

[3] J. L. Kim, G. Kollias, A. Kalev, K. X Wei, and A. Kyrillidis. *arXiv:2104.07006.*