GANIT

Available online at https://ganitjournal.bdmathsociety.org/

GANIT: Journal of Bangladesh Mathematical Society

GANIT *J. Bangladesh Math. Soc.* 43.2 (2023) 65–74
DOI: https://doi.org/10.3329/ganit.v43i2.70799.

Quantum State Reconstruction Through Online Shadow Tomography: Theoretical Framework and Simulation Results

Rashedul Islam Seum, Md. Nirab Hossain *, Jenia Fardousi Koly, Md. Showkat Ali, and Nur Mohammad Salem

Department of Applied Mathematics, Faculty of science, University of Dhaka, Dhaka 1000, Bangladesh

ABSTRACT

The purpose of this research work is to learn the quantum states in an ideal environments analytically, computationally, and graphically. The analysis starts with the learning of quantum states in identity channels with the help of the Regularized Follow the Leader (RFTL) algorithm. Our machine will try to learn the states based on the previous information, which is called the online learning model. The objective of this problem is to minimize regret by utilizing a learning algorithm that successively anticipates quantum states through observed measurements and losses. We have to produce many copies of quantum state ρ to perform analysis on them, which indicates the use of the shadow tomography approach in an ideal situation. Our goal is to learn the shadow of the state ρ by using a series of measurement operators that have two outcomes in nature. Aaronson et al. [1] developed an online setting for a non-realizable case, where the maximum possible loss is $O(\sqrt{Tn})$ for the best possible state up to $T-$measurements . It is noteworthy that this outcome is an extension of the Aaronson PAC-like findings [2].

Received: September 25, 2023 **Accepted:** December 29, 2023 **Published Online:** December 31, 2023

Keywords: Quantum State Tomography; Shadow Tomography; Online Learning; Quantum Machine Learning; Learnability of Quantum States; Regularized Follow The Leader Algorithm

1 Introduction

In the early 1980s, machine learning became an active utility of research in theoretical computer science. Probably Approximately Correct (PAC) model was developed by Leslie Valiant [3], it plays a vital role to popularize machine learning and is based on the previous model that has been used in statistics and was developed by Vapnik and others. Leslie Valiant, however, imposes a new concept of computational complexity in this model. The mathematical reliability definition attributes an efficient framework to learn the target concepts. Deutsch's definition of a Universal Turing Machine [4] subsumed quantum computing into digital realm, and Peter Shor's work on his factoring algorithm [5] pushed it one step further.

Quantum computing fundamentally aims to learn the description of an unknown quantum state ρ. Traditionally, this problem is known as *quantum state tomography*, where the learner attempts to learn the quantum state ρ by performing arbitrary measurements and come up with an approximate state σ such that both are arbitrarily

*Corresponding author. *E-mail address:* nirab@du.ac.bd

close in trace distance. Scott Aaronson [2] presented this problem in the Probably Approximate Correct (PAC) framework, which has had a significant impact on quantum computing. In this article, instead of learning ρ under an absolute metric (*such as trace distance*), we only wish to learn ρ with respect to a pre-specified underlying distribution on the measurements.

Scoot Aaronson in a later work [1] extended the PAC setting into an online setting where we answer the question: is it true that given a sequence $E_1, E_2, \cdots$ of yes/no measurements, where each E_t is followed shortly by a $d_{\mathrm{Tr}}(E_t\rho)$, there is a way to anticipate $d_{\mathrm{Tr}}(E_t\rho)$ by guesses $y_t \in [0,1]$ in such a way that $|y_t - \mathrm{Tr}(E_t\rho)| > \epsilon$ at most $O(n)$ times, given $\epsilon > 0$ and n is the number of qubits? It turns out that the answer is yes, and we can accomplish this efficiently. Note that instead of learning the full quantum state, we are learning the actions of a quantum state subject to a measurement which are referred to as the shadows of the quantum state.

2 Background and Literature Review

In this section, we will discuss three learning models: the Exact learning model, *PAC* learning model and Agnostic learning model introduced by Angluin, Valiant, and Haussler and Kearns et al., respectively. The three learning models will be described both in classical and quantum settings. We will introduce the concept of *Quantum Example Oracle* (QPEX) in all of those models. Most importantly, the models will be in an oracular setting. That is, we shall consider a concept class $\mathcal{C}$ which can be considered as a set of functions as $c : \{0,1\}^n \mapsto \{0,1\}$ or $c : [N] \mapsto \{0,1\}$, where N is a sepcified bit string.

2.1 The Query Model of Computation

The main objective of this model is to compute a Boolean function $\{0,1\}^n \mapsto \{0,1\}$. Here we encode the bits of X instead of knowing it explicitly, i.e., the oracle will return x_i for some given $i \in [N]$. The cost of such an algorithm is equal to the number of queries to the oracle or number of times an oracle is accessed.
Formally a quantum query is equivalent to constructing the following unitary over $n+1$ qubits:

$$O_x : |i, b\rangle \mapsto |i, b \oplus x_i\rangle$$

Where $i \in 0, 1, \cdots N-1$, and $N = 2^n$, $b \in [0,1]$.

2.2 Exact Learning

Classical Exact Learning: A learner $\mathcal{A}$ is trying to learn something for a target concept $c \in \mathcal{C}$ that has given access to some specific membership oracle MQ(c). MQ(c) gives back the label $c(x)$ for the input $x \in \{0,1\}^n$. A learning algorithm $\mathcal{A}$ is said to be an exact learning algorithm if the following condition is satisfied:
For all $c(x)$ there is a hypothesis output $h(x)$ such that

$$Pr(h(x) = c(x)) \geq \tfrac{2}{3}, \qquad \forall x \in \{0,1\}^n \tag{2.1}$$

As we can see, the full procedure is based on the identification of some actual oracles $c \in \mathcal{C}$ which is why this is also known as "Oracle Identification". The query complexity of the learner $\mathcal{A}$ is given by for all $c \in \mathcal{C}$ and internal randomness x the maximum number of accesses to the oracle MQ(c).

Quantum Exact Learning: Analogue to classical, in quantum setting, the learner will have access to Quantum Membership Oracle QMC(c) instead of Membership Oracle MQ(c) in classical ones, that is the mapping:

$$QMC(c) : |x, b\rangle \mapsto |x, b \oplus c(x)\rangle, \qquad x \in \{0,1\}^n,\ b \in \{0,1\}$$

The matrices are equivalent in both setups. The *query complexity* is defined as the minimum number of queries over all $\mathcal{C}$. By using the truth table of $N = 2^n$-bit, each concept $c \in \mathcal{C}$ can also be defined in an alternative way. If we can maximize query complexity over all $\mathcal{C} \subseteq \{0,1\}^N$ such that $|\mathcal{C}| = M$ then we can define the (N, M)-query complexity.

2.3 Probably Approximate Correct (PAC) Learning

Classical PAC Learning: A learner $\mathcal{A}$ is trying to learn from a target concept $c \in \mathcal{C}$ and a distribution $D : \{0,1\}^n \mapsto [0,1]$ which together make a *random example oracle* $PEX(c, D)$. Note that $PEX(c, D)$ will give $(x, c(x))$, where x is taken from the unknown distribution D. We can define (ϵ, δ)-PAC learning for algorithm $\mathcal{A}$ if-

$$Pr_{x\in D}([h(x) \neq c(x)] \leq \epsilon) \geq 1 - \delta$$

Here, the learner is independent of the output hypothesis, which is not a member of the target concept. If we can do so, then it is called proper PAC learning. The sample complexity for $\mathcal{A}$ is defined as the maximum calls to $PEX(c, D)$. The (ϵ, δ)-PAC sample complexity is defined as the minimum over all (ϵ, δ)-PAC.

Quantum PAC Learning: Here the learner [6] is given access to *Quantum Example Oracle* $QPEX(c, D)$ and which uses superposition to give a quantum example.

$$\sum_{x\in\{0,1\}} \sqrt{D(x)}|x, c(x)\rangle$$

A learner here performs a POVM measurement on the state, and for each outcome, the learner assigns a hypothesis given that he has multiple copies of the quantum example. We can also define here (ϵ, δ) quantum PAC and the sample complexity is the minimum number of queries over all $QPEX(c, D)$.

2.4 Quantum State Tomography and Learning Quantum States

State tomography is a process where an apparatus is capable of producing many copies of the quantum state, and by performing suitable measurements, we can get a description of the system. This description will give the accuracy of the apparatus. These experiments involve sending quantum states over long distances. But how does the receiving party verify that the state that they receive is the same as the one that was sent by the other party? This is a problem of significant practical importance and is precisely what tomography aims to solve.

The question is, to learn the state ρ density matrix of dimension $(2^n \times 2^n)$ at ϵ trace distance, how many single copies of the state are needed? Depending on the assumptions taken on the ρ it depends on, if there is no assumption, then measurement numbers grow exponentially with n. Even if we take joint measurements, the number of measurements also grows exponentially [7, 8].

More generally, *state tomography* is the process of learning the given unknown mixed state $\rho \in \mathbb{C}^{d\times d}$. But it is restricted because it is then possible to recreate the state ρ if all of the information regarding it is given, which also indicates that we are violating *no-cloning* theorem. It turns out that we can't know a lot of detail about the mixed state.

Let's say the state ρ is measured in basis $\{|v_1\rangle, |v_2\rangle, \cdots |v_m\rangle\}$ and after measurement the state is $|v_i\rangle$. The probability $\langle v_i|\rho|v_i\rangle$ is the only quantity we can measure, and we are guessing that the probability amplitude is somewhat large. This statement provides limited but valuable insights regarding the state ρ. But as we have observed, the state is now broken. So the adjustment will be: we need a lot of copies so that we can perform measurements remotely. Suppose there is an apparatus that can make n copies of the state ρ. As the state ρ is too restrictive, so our second adjustment: with high probability, the output $\tilde{\rho} \in \mathbb{C}^{d\times d}$ is the approximation of the original state ρ.

In generally, a Hermitian matrix of $2^n \times 2^n$ describe a $n-$qubit quantum states which have essentially need to be trace of unity, which leads to the fact that it needs $\sim 2^{2^n}$ real parameters. Aaronson has shown in his paper [2] how to effectively learn those states.

Definition 1. *In quantum state tomography problem, we want to estimate the mixed state ρ which has dimension d given n copies of the state $\rho^{\otimes n}$ where $n, d \in \mathbb{R}$. Specifically, this process has the following input and output:*

Input: With an error margin $\epsilon \geq 0$, a tensor product of power n, which is $\rho^{\otimes n}$ of the mixed state ρ.
Output: Approximate state $\tilde{\rho} \in \mathbb{C}^{d\times d}$ such that the trace distance is within the error margin with high probability,

$$d_{\mathrm{Tr}}(\rho, \tilde{\rho}) \leq \epsilon$$

In general, a Hermitian matrix of dimension $2^n \times 2^n$ represent $n-$qubit system has trace unity and has 2^{2^n} real parameters. Aaronson studied in his paper how to effectively learn those states [2]. We studied the case with the restriction of two outcome measurements, E_i and $\mathbb{I} - E_i$, such that $\mathrm{Tr}(E_i\rho)$ denoted the probability of each outcome.
Aaronson stated a very important PAC-learning result [2] which is the fundamental point for quantum state tomography. The result is that, from $O(n)$ measurement outcomes, which are randomly selected i.i.d. from an unknown distribution D on the complete set of two-outcome measurements, it is possible to generate the state $\tilde{\rho}$ such that it has roughly the same expectation as the state ρ. Formally, this can be written as:

Theorem 1. *For the parameters $\gamma,\ \delta,\ \epsilon > 0\ \ \exists$ learner has the following: For the set of two outcome measurements E_i which are taken i.i.d. from distribution D and for given $T = n \cdot poly(\frac{1}{\epsilon}, \frac{1}{\gamma}, \log(\delta))$ and for the measurement results $(E_1, b_1), (E_2, b_2), \cdots, (E_n, b_n)$ where b_i is a bit with $Pr(b_i = 1) = Tr(E_i\rho)$ with the probability $\geq 1 - \delta$, a classical description $\tilde{\rho}$ is produced by the learner as:*

$$Pr_{E\sim D}\left[Tr(E\tilde{\rho}) - Tr(E\rho) > \gamma\right] \leq \epsilon$$

The proof of this theorem uses the results of Anthony and Bartlett [9] and Bartlett and Long [10] on γ shattering dimension, which uses a Boolean function and VC dimensions as follows:

Proof. Let, for a set $\mathcal{E}$ there be a class of function $f : \mathcal{E} \mapsto [0, 1]$. If there exists $\alpha_1, \alpha_2, \cdots, \alpha_d$ for some set of measurements $S = \{E_1, E_2, \cdots, E_d\}$ such that there exists $f \in \mathcal{C}$ for some $Z \subseteq [d]$:
Case-1: If $i \in Z$ then $f(E_i) \geq \alpha_i + \gamma$
Case-2: If $i \notin Z$ then $f(E_i) \leq \alpha_i - \gamma$
The largest set S that is shattered by $\mathcal{C}$ is known as the $\gamma-$fat shattering dimension of $\mathcal{C}$.

This indicates that with probability α_i with respect to the measurement E_i, which is recovered from a bit z_i with a $\gamma-$fat shattering there exists ρ_z for each string $z \in \{0, 1\}^d$. The encodings from the classical state to the quantum state $z \mapsto \rho_z$ are called quantum random access codes. Aaronson showed that for known bounds [11] of such codes, $d = O(\frac{b}{\gamma^2})$. □

2.5 Online Learning of Quantum States

According to Aaronson the measurement operators are simply drawn from distribution D leads to a major drawback; moreover it is the same distribution from where the training samples are picked for testing the learners performance. In the context of learning theorist this leads to an inefficient situation as they failed for the evolving environments. In order to overcome this, we need to focus on the online learning setting which is based on the learning theorem by Aaronson. Mathematically, is it true that, given a sequence $E_1, E_2, \cdots$ of yes/no measurements, where each E_t is followed shortly by a $d_{\mathrm{Tr}}(E_t\rho)$, there is a way to anticipate $d_{\mathrm{Tr}}(E_t\rho)$ by guesses $y_t \in [0, 1]$ in such a way such that $|y_t - \mathrm{Tr}(E_t\rho)| > \epsilon$ at most $O(n)$ times, given $\epsilon > 0$ and n is the number of qubits? This paper aims to address this question.

The query model of computation is generalized when we try to find out answer of our above question which eventually leads to online setting and regret minimization. We will use the main results from paper [1].

Theorem 2. *Let $E_1, E_2, \cdots$ be a sequence of two outcome measurements, and let ρ be an $n-$qubit mixed state. Here the measurement operators are presented to the learner one by one, where each of them is associated with a value $b_t \in [0, 1]$ such that $|Tr(E_t\rho) - b_t| < \frac{\epsilon}{3}$. Then an explicit strategy exists for generating hypotheses states $\tilde{\rho}_1, \tilde{\rho}_2, \cdots$ such that $|Tr(E_t\tilde{\rho}_t) - Tr(E_t\rho)|$ for at most $O(\frac{n}{\epsilon^2})$ values of t.*

Proof. A sketch of the proof is as follows:

1. We will start from a maximally mixed state, $\tilde{\rho}_1 = \frac{\mathbb{I}}{2^n}$.
2. Now for each time step $t = 1, 2, 3, \cdots$ we have the following:
   (a) Predict $\tilde{\rho}_t$. Due to the measurement E_t, the learner has to face loss $l_t : \mathbb{R} \mapsto \mathbb{R}$. The loss function $l_t(\mathrm{Tr}(E_t\psi))$ is surely $L-$Lipschitz and convex. Let $l'_t(x)$ denotes the subderivative with respect to x.
   (b) Let's define $\nabla_t \coloneqq l'_t(\mathrm{Tr}(E_t\tilde{\rho}_t))E_t$. The decision will be updated following the RFTL algorithm with the von Neumann regularizer as follows:

$$\tilde{\rho}_{t+1} = \arg\min_{\psi} \left\{ \eta \sum_{s=1}^{t} \mathrm{Tr}(\nabla_s\psi) + \sum_{i=1}^{2^n} \lambda_i(\psi) \log \lambda_i \right\}$$

(c) The iteration's progression is as follows: Let at t^{th} iteration the hypothesis be $\tilde{\rho}_t$. The current hypothesis will run if the value of the loss function is $l > \frac{2\epsilon}{3}$. This algorithm makes at most $O(\frac{n}{\epsilon^2})$ such updates.

□

There is an alternative approach to updating the hypothesis, which is called the Matrix Multiplicative Weights (MMW) discussed in [12] and the new update rule would be:

$$\tilde{\rho}_{t+1} = \frac{\exp(-\frac{\eta}{L}\sum_{\tau=1}^{t}\nabla_\tau)}{\text{Tr}(\exp(-\frac{\eta}{L}\sum_{\tau=1}^{t}\nabla_\tau))}$$

The regret of theorem 2 can be found by setting $\eta = \sqrt{\frac{(\log\ 2)n}{2TL^2}}$ and is bounded by $2L\sqrt{(2\log 2)Tn}$, on the other hand, while using MMW updates by using $\eta = \sqrt{\frac{(\log 2)n}{4T}}$ the regret is bounded by $2L\sqrt{(\log 2)Tn}$, where the regret R_T is defined by:

$$R_T \coloneqq \sum_{t=1}^{T} l_t(\text{Tr}(E_t\tilde{\rho}_t)) - \min_{\psi\in C_n}\sum_{t=1}^{T} l_t(\text{Tr}(E_t\psi))$$

Therefore in general for an $n-$qubit system it can be shown that if $E_1, E_2, \cdots$ be a sequence of two outcome measurements and $l_1, l_2, \cdots$ be the corresponding convex and $L-$Lipschitz loss function then there Aaronson et. al. [1] guarantees that there is a learning algorithm with regret $R_T = O(L\sqrt{Tn})$.

3 Computational Simulations

In this final section, we will visualize our simulations for both online learning of quantum states and noisy online learning of quantum states. In the case of visualization, *fidelity* is crucial to understanding the topic clearly.

Fidelity: *The fidelity between two quantum states, ϕ and $\tilde{\phi}$ is denoted by F and is defined as*

$$F(\phi, \tilde{\phi}) = Tr\left[\sqrt{\sqrt{\phi}\tilde{\phi}\sqrt{\phi}}\right]$$

It represents the quality of learning between the approximate and true quantum states. The value of F has the range $[0, 1]$, where 0 indicates either the states are orthogonal or there exists dissimilarity, and 1 indicates that the two states are completely identical.

Fidelity between two pure states $\phi = |\psi_1\rangle\langle\psi_1|$ and $\tilde{\phi} = |\psi_2\rangle\langle\psi_2|$ reduces to its square overlap:

$$F(\phi, \tilde{\phi}) = |\langle\psi_1|\psi_2\rangle|^2$$

3.1 Simulations of Online Learning of Quantum States

Using the RFTL algorithm for learning quantum states, simulations were performed using Python programming language. The outcomes of the simulations are depicted in graphical form.

Initially, we will demonstrate the loss and fidelity values for various iterations and learning rates. Subsequently, we will visually represent these values through graphs and draw conclusions based on our analysis.

In the simulation, we are trying to learn the target state: $\rho = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Case I: Simulation for online learning of quantum state for $\eta = 0.5$ and $T = 10$:

Table 3.1: Loss and Fidelity for $\eta = 0.5$ and $T = 10$.

Iteration	Loss	Fidelity
$t = 1$	0.2475	0.7071
$t = 2$	0.9851	0.0000
$t = 3$	0.4060	0.6065
$t = 4$	0.4602	0.5739
$t = 5$	0.1413	0.5805
$t = 6$	0.1160	0.6065
$t = 7$	0.3943	0.6065
$t = 8$	0.1427	0.5810
$t = 9$	0.4104	0.6065
$t = 10$	0.1284	0.5819

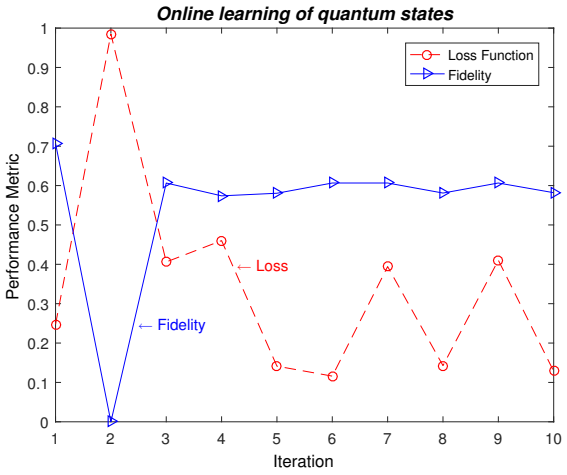

Figure 3.1: Learning Quantum States with $\eta = 0.5$ and $T = 10$.

It is expected that the identity channel will suffer a higher loss initially, but with time, the machine will learn to estimate the states' results with increasing fidelity. Now we analyze the results by increasing the value of the time step.

Case II: Simulation for online learning of quantum state for $\eta = 0.5$ and $T = 20$:

Table 3.2: Loss and Fidelity for $\eta = 0.5$ and $T = 20$.

Iteration	Loss	Fidelity
$t = 1$	0.2597	0.7071
$t = 2$	1.0026	0.0000
$t = 3$	0.1351	0.6065
$t = 4$	0.3932	0.6066
$t = 5$	0.4440	0.5775
$t = 6$	0.1289	0.5818
$t = 7$	0.4030	0.6065
$t = 8$	0.1379	0.5810
$t = 9$	0.3952	0.6065
$t = 10$	0.4348	0.5819
$t = 11$	0.4408	0.5841
$t = 12$	0.1344	0.5842
$t = 13$	0.1096	0.6065
$t = 14$	0.3922	0.6065
$t = 15$	0.1350	0.5829
$t = 16$	0.4086	0.6066
$t = 17$	0.1360	0.5832
$t = 18$	0.1148	0.6065
$t = 19$	0.4099	0.6065
$t = 20$	0.1302	0.5834

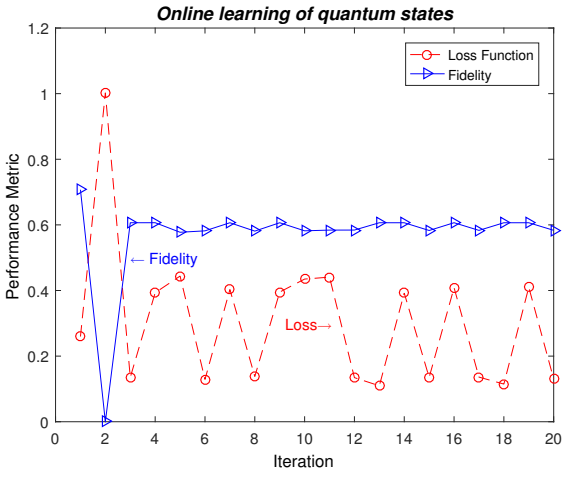

Figure 3.2: Learning Quantum States with $\eta = 0.5$ and $T = 20$.

Again, with respect to the increasing time step, the identity channel will suffer a higher loss initially, but eventually, the machine will learn to estimate the state's results with increasing fidelity. And most importantly, the each of the graphs is moving towards its equilibrium.

Case III: Simulation for online learning of quantum state for $\eta = 0.5$ and $T = 30$:

Table 3.3: Loss and Fidelity for $\eta = 0.5$ and $T = 30$.

Iteration	Loss	Fidelity
$t = 1$	0.2591	0.7071
$t = 2$	1.0001	0.0014
$t = 3$	0.3928	0.6065
$t = 4$	0.1335	0.5584
$t = 5$	0.3876	0.6065
$t = 6$	0.1286	0.5661
$t = 7$	0.3876	0.6065
$t = 8$	0.1366	0.5687
$t = 9$	0.1071	0.6065
$t = 10$	0.1071	0.6065
$t = 11$	0.4049	0.6066
$t = 12$	0.1285	0.5708
$t = 13$	0.1045	0.6065
$t = 14$	0.1063	0.6065
$t = 15$	0.3985	0.6065
$t = 16$	0.4497	0.5717
$t = 17$	0.4523	0.5756
$t = 18$	0.1329	0.5753
$t = 19$	0.3958	0.6065
$t = 20$	0.1391	0.5722
$t = 21$	0.4037	0.6065
$t = 22$	0.1323	0.5724
$t = 23$	0.3897	0.6065
$t = 24$	0.1326	0.5725
$t = 25$	0.1037	0.6065
$t = 26$	0.1071	0.6065
$t = 27$	0.1083	0.6065
$t = 28$	0.4039	0.6065
$t = 29$	0.4504	0.5728
$t = 30$	0.4476	0.5764

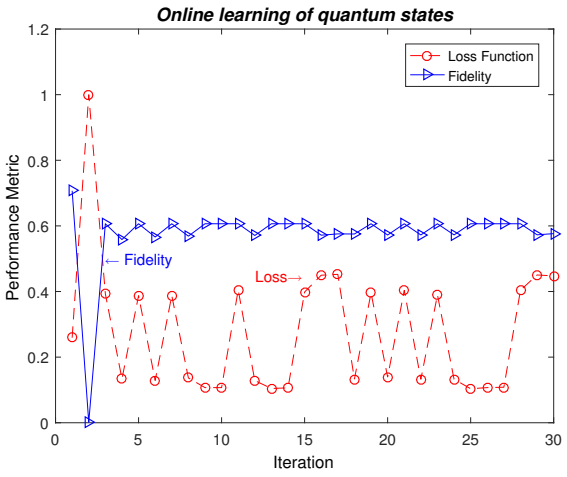

Figure 3.3: Learning Quantum States with $\eta = 0.5$ and $T = 30$.

Finally, using theoretical and graphical analysis, we can conclude that the machine is exhibiting proficient learning capabilities in relation to the identity channel. By increasing the time step and learning rate with some

specific tolerance, fidelity increases and loss decreases. After a long time, their parallel movement indicate the effectiveness of the RFTL algorithm.

4 Conclusion

This article validates that the RFTL algorithm is applied for learning quantum states under an ideal environment. It shows the efficacy of RFTL algorithm as the number of iterations increases. Simulation shows that loss and fidelity curves are moving to an equilibrium with gradual time step progression. This convergence of loss and fidelity over time serves as a testament to the effectiveness of this approach, offering exciting prospects for further research and practical applications in Quantum Machine Learning. In future, further generalization in this setting to a noisy channel can be established, where the state will change at every round with respect to a noise parameter.

References

[1] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, pp. 124019, 2019.

[2] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, pp. 3089–3114, 2007.

[3] L. Valiant. A theory of the learnable. *Communications of the ACM*, vol. 27, pp. 1134–1142, 1984.

[4] David Deutsch. Quantum theory, the Church–Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 400, pp. 97–117, 1985.

[5] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, vol. 25, pp. 1484—1509, 1997.

[6] N. H. Bshouty and J. C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, vol. 28, pp. 1136—1153, 1999.

[7] Ryan O'Donnell and John Wright. Efficient quantum tomography, pp. 441–459, 2015.

[8] Jeongwan Haah and Aram W. Harrow and Zhengfeng Ji and Xiaodi Wu and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 2017.

[9] Anthony, Martin and Bartlett, Peter. Dense Quantum Coding and Quantum Finite Automata. Efficient quantum tomography. *The European Physical Journal C*, vol. 904, pp. 211–221, 2006.

[10] Peter L. Bartlett and Philip M. Long. Prediction, Learning, Uniform Convergence, and Scale-Sensitive Dimensions. *Journal of Computer and System Sciences*, vol. 56, pp. 174–190, 1998.

[11] Ambainis, Andris and Nayak, Ashwin and Ta-Shma, Amnon and Vazirani, Umesh. Dense Quantum Coding and Quantum Finite Automata. Efficient quantum tomography. *J. ACM*, vol. 49, pp. 496–511, 2002.

[12] Koji Tsuda, Gunnar R¨atsch, and Manfred K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *Journal of machine learning research*, vol. 6, pp. 995—1018, 2005.

[13] Mark M. Wilde. Quantum Information Theory. *Cambridge University Press*, 2016.

[14] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum principal component analysis. *Nature physics*, vol. 10, pp. 631—633, 2013.

[15] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *Nature physics*, 2012.

[16] P. Rebentrost, M. Mohseni, and S. Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, vol. 113, 2013.

[17] M. Schuld, I. Sinayskiy, and F. Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, vol. 66, pp. 172—185, 2015.

[18] A. Ambainis, A. Nayak, A. Ta-Shma, and U. V. Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, vol. 49, pp. 496—511, 2002.