Online Quantum Learning

Aayush Karan

1 Introduction

Quantum state tomography concerns the task of learning an unknown quantum state ρ from a series of measurements conducted on copies of the state generated by some black box apparatus. Of primary interest is the *sample complexity*, or the minimum number of copies required to learn the unknown state ρ sufficiently well. The choice of setting that makes "sufficiently well" precise can drastically affect the sample complexity. For example, if we want to learn the density matrix representation of ρ up to small error in trace norm (see Definition 1.5) distance, then for independently chosen (nonadaptive) measurements, the sample complexity is necessarily at least cubic in the dimension of ρ [3]. On the other hand, if we want to accurately predict expectation values of randomly drawn binary measurements on ρ, it turns out that the sample complexity is logarithmic in the dimension of ρ [2].

For this paper, we are concerned with the online learning setting. In *online learning*, a learner receives a sequence of data $x_1, x_2, \cdots$ over a period of time; at timestep t, a prediction is made and environmental feedback is returned. In this setting, as opposed to directly limiting some notion of sample complexity, we would like to limit the number of *mistakes*, which occur when the feedback on a prediction violates some error threshold. Online learning algorithms with a finite (non-time dependent) limit on mistakes is called a *mistake-bounded learning algorithm*. For learning quantum states online, mistakes correspond to high error in estimating expectation values of measurement predictions. As we will see shortly, we can actually design a mistake-bounded learning algorithm with mistake bound that is logarithmic in the dimension of the quantum state [1]. Before formalizing online quantum learning, we introduce some notation and prerequisite mathematical knowledge.

1.1 Preliminaries

1.1.1 Positive Semidefinite Matrices

In this section, let us fix M, N, and P to be arbitrary complex Hermitian matrices of dimension d. This choice is to avoid description redundancy in what follows.

Definition 1.1. M is said to be *positive semidefinite (PSD)* if all its (real) eigenvalues are nonnegative. A shorthand notation to designate M as PSD is to write $M \succeq 0$. For given M, N, we write $M \succeq N \iff M - N \succeq 0$.

Definition 1.2. The *spectrum* of M, written as $\mathrm{Spec}(M)$, is the multiset of all (real) eigenvalues of M.

Lemma 1.3. *Suppose we have* $M, N \succeq 0$. *Then* $\mathrm{Tr}(MN) \geq 0$.

Proof. Due to the well-known Cholesky decomposition, we can write $N = SS^\dagger$, where S is positive semidefinite and Hermitian. Notice $S^\dagger M S \succeq 0$, since for any vector $v \in \mathbb{C}^d$, $v^\dagger S^\dagger M S v = (Sv)^\dagger M (Sv) \geq 0$ as $M \succeq 0$. Hence, $\mathrm{Tr}(MN) = \mathrm{Tr}(S^\dagger M S) \geq 0$, as desired. □

Corollary 1.4. *Suppose $P \succeq 0$ and $M \succeq N$. Then* $\mathrm{Tr}(MP) \geq \mathrm{Tr}(NP)$.

Proof. Apply Lemma 1.3 to P and $M - N$. $\square$

Definition 1.5. The *trace norm* of M is given by $||M||_{\mathrm{Tr}} = \mathrm{Tr}\left(\sqrt{MM^\dagger}\right)$; in other words, the trace norm is sum of the absolute values of the eigenvalues of M.

Lemma 1.6. *If the trace norm of M satisfies $||M||_{Tr} \leq 1$, then $M \preceq \mathbb{1}$, where $\mathbb{1}$ denotes the identity matrix.*

Proof. Since the trace norm of M is the sum of the absolute values of its eigenvalues, it follows that all the eigenvalues of M have absolute value less than 1. Now $\mathbb{1}$ and M are simultaneously diagonalizable, so the eigenvalues of $\mathbb{1} - M$ are of the form $1 - \lambda$ for $\lambda \in \mathrm{Spec}(M)$. But $|\lambda| \leq 1 \implies 1 - \lambda \geq 0$ for all such λ, so it follows that $M \preceq \mathbb{1}$, as desired. $\square$

Lemma 1.7. *Suppose the trace norm of M satisfies $||M||_{Tr} \leq 1$. Then*

$$\exp(-M) \preceq \mathbb{1} - M + M^2.$$

Proof. Let D be the diagonalization of M, so that $M = UDU^\dagger$ for some unitary matrix U. Notice that $\exp(-D) \preceq \mathbb{1} - D + D^2$ is equivalent to proving the inequality on the real eigenvalues λ so that $\exp(-\lambda) \leq 1 - \lambda + \lambda^2$ for $|\lambda| \leq 1$. The latter holds (see Appendix) whence the former does as well, so reconjugating both sides of the relation on D by U and $U^\dagger$ finishes the proof of the lemma, since $U \exp(-D) U^\dagger = \exp(-UDU^\dagger) = \exp(-M)$ and $UD^2U^\dagger = MUDU^\dagger = M^2$. $\square$

Proposition 1.8 (Golden-Thompson Inequality; see [7])**.** *The following inequality on M, N holds:*

$$\mathrm{Tr}(\exp(M + N)) \leq \mathrm{Tr}(\exp(M)\exp(N)).$$

1.1.2 Quantum States and Measurements

Definition 1.9. In the density matrix formalism [8], a *d-dimensional quantum state* ρ is a PSD matrix of dimension d with unit trace. Let $\mathbb{D}(d)$ denote the set of all such quantum states.

Definition 1.10. A *binary measurement operator* of state ρ is a Hermitian matrix E such that $0 \preceq E \preceq \mathbb{1}$ (alternatively, its eigenvalues lie in the interval $[0, 1]$). This is a generalization [8] of a projective measurement (whose eigenvalues lie in $\{0, 1\}$), but the interpretation remains that the probability of a positive result upon measuring ρ is given by $\mathrm{Tr}(E\rho)$ and a negative result $1 - \mathrm{Tr}(E\rho)$.

1.2 Framework

With the basic preliminaries in place, we can now formalize the quantum online learning setting. For the remainder of this paper, fix a positive integer d and an arbitrary unknown d-dimensional quantum state $\rho \in \mathbb{D}(d)$. The online learner maintains a sequence of estimates $\{\omega_t\}_{t \geq 1}$ of ρ over time, with all $\omega_t \in \mathbb{D}(d)$, the goal being to predict any measurement expectation over the true state ρ sufficiently accurately. Fix error parameter $\varepsilon \in (0, 1)$. For times $t = 1, 2, 3, \cdots$, the learner receives binary measurement operator E_t from the environment, with the prediction being $\mathrm{Tr}(E_t \omega_t)$, which can be calculated since both E_t and ω_t are known. The environment responds by outputting an approximation b_t of $\mathrm{Tr}(E_t \rho)$ such that

$$|b_t - \mathrm{Tr}(E_t \rho)| \leq \frac{\varepsilon}{3}. \tag{1}$$

The means by which the approximation b_t is obtained is typically relegated to an oracle [1]; however, one simple methodology is to simply take the empirical average of several copies of ρ.

A prediction is a *mistake* if $|\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\rho)| > \varepsilon$, and we would like to bound the number of mistakes our learning scheme makes so that after some time T, our estimates $\{\omega_t\}_{t\geq T}$ are guaranteed to satisfy $|\mathrm{Tr}(E\omega_t) - \mathrm{Tr}(E\rho)| \leq \varepsilon$ for any binary measurement operator E. The key result this paper focuses on is the following mistake bound, due to [1].

Theorem 1.11. *Let ρ be a d-dimensional quantum state. Then for any sequence of binary measurement operators $E_1, E_2, \cdots$ in an online learning setting, there exists an algorithm constructing estimates sequence $\{\omega_t\}_{t\geq 1}$ making at most $O\left(\frac{\log(d)}{\varepsilon^2}\right)$ mistakes, given environmental feedback sequence $\{b_t\}_{t\geq 1}$ satisfying $|b_t - Tr(E_t\rho)| \leq \frac{\varepsilon}{3}$ for all t.*

The subsequent exposition and proof of this theorem roughly follows [1] and [3].

2 The Matrix Multiplicative Weights Algorithm

The Matrix Multiplicative Weights (MMW) algorithm [6] highlights the powerful technique of adaptive weighting methods in online learning which have proven remarkable results in classical computational learning theory [4], [5]. We begin with a general performance bound on the algorithm.

The MMW algorithm works as follows. Fix update hyperparameter $\beta \in (0, 1)$. Suppose further that for each timestep $t = 1, 2, \cdots$, given the learner prediction and environmental feedback, one can define a *loss matrix* L_t, where L_t is a d-dimensional Hermitian matrix such that $||L_t||_{\mathrm{Tr}} \leq 1$. Initialize weight matrix $W_1 = \mathbb{1}$ to the identity matrix. Then at time t, we have the following update rule:

- Update the weight matrix to
$$W_t := \exp\left(-\beta \sum_{i=1}^{t-1} L_i\right). \tag{2}$$
- Compute the estimate $\omega_t := \frac{W_t}{\mathrm{Tr}(W_t)}$.
- Compute loss matrix L_t given ω_t, binary measurement operator E_t, and feedback b_t.

Lemma 2.1. *The weight matrix W_t is positive semidefinite for all times t.*

Proof. Fix time t. Observe that $\sum_{i=1}^{t-1} L_i$ is Hermitian as it is the sum of individual Hermitian matrices. If $\mathrm{Spec}\left(\sum_{i=1}^{t-1} L_i\right) = \{\lambda_i\}_{i=1}^{d}$, then $\mathrm{Spec}(W_t) = \{\exp(-\beta\lambda_i)\}_{i=1}^{d}$; in other words, W_t has all eigenvalues nonnegative. It follows that W_t is PSD for all times t, as desired. □

Corollary 2.2. *For all times t, the estimate ω_t is contained in $\mathbb{D}(d)$ and is hence a valid d-dimensional quantum state.*

Proof. Fix time t. We can quickly confirm that $\mathrm{Tr}(\omega_t) = \frac{\mathrm{Tr}(W_t)}{\mathrm{Tr}(W_t)} = 1$, so ω_t has unit trace. From Lemma 2.1 $W_t \succeq 0$, so $\frac{W_t}{\mathrm{Tr}(W_t)} \succeq 0$. Hence, ω_t is PSD with unit trace $\implies \omega_t \in \mathbb{D}(d)$ for all times t, as desired. □

Theorem 2.3. (See [3]) *For any time T, the estimates $\{\omega_t\}_{t\leq T}$ of the MMW algorithm satisfy*

$$\sum_{t=1}^{T} \mathrm{Tr}(L_t\omega_t) \leq \lambda_{\min}\left(\sum_{t=1}^{T} L_t\right) + \beta\sum_{t=1}^{T}\mathrm{Tr}(L_t^2\omega_t) + \frac{\log(d)}{\beta}, \tag{3}$$

where λ_{min} is a function that outputs the minimum eigenvalue of a Hermitian matrix.

Proof. We prove this inequality by providing upper and lower bounds on the quantity $\mathrm{Tr}(W_{T+1}) = \mathrm{Tr}\Big(\exp\left(-\beta\sum_{i=1}^{T} L_i\right)\Big)$. For the lower bound, since $\sum_{i=1}^{T} L_i$ is Hermitian, let $\mathrm{Spec}\Big(\sum_{i=1}^{T} L_i\Big) = \{\lambda_i\}_{i=1}^{d}$. Then clearly

$$\mathrm{Tr}\left(\exp\left(-\beta\sum_{i=1}^{T} L_i\right)\right) = \sum_{i=1}^{d}\exp\left(-\beta\lambda_i\right) \geq \exp\left(-\beta\lambda_{\min}\left(\sum_{t=1}^{T} L_t\right)\right), \tag{4}$$

where the former equality is due to the fact that $\mathrm{Tr}(\exp M) = \sum_{\lambda\in\mathrm{Spec}(M)}\exp(\lambda)$. The latter inequality holds since $\exp(-\beta\cdot\lambda_i)\geq 0$ for all i.

For the upper bound, notice that for any t,

$$\mathrm{Tr}\left(W_{t+1}\right) = \mathrm{Tr}\left(\exp\left(-\beta\sum_{i=1}^{t} L_i\right)\right) = \mathrm{Tr}\left(\exp\left(-\beta L_t - \beta\sum_{i=1}^{t-1} L_i\right)\right) \leq \mathrm{Tr}\left(\exp\left(-\beta L_t\right)W_t\right), \tag{5}$$

using Lemma 1.8. Now, since $||\beta L_t||_{\mathrm{Tr}} \leq ||L_t||_{\mathrm{Tr}} \leq 1$, from Lemma 1.7 we have that

$$\exp\left(-\beta L_t\right) \preceq \mathbb{1} - \beta L_t + \beta^2 L_t^2.$$

From Lemma 2.1, we know that $W_t \succeq 0$, so from Corollary 1.4, we have

$$\mathrm{Tr}\left(\exp\left(-\beta L_t\right)W_t\right) \leq \mathrm{Tr}\left(W_t - \beta L_t W_t + \beta^2 L_t^2 W_t\right) = \mathrm{Tr}(W_t)\left(1 - \beta\mathrm{Tr}\left(L_t\omega_t\right) + \beta^2\mathrm{Tr}\left(L_t^2\omega_t\right)\right). \tag{6}$$

By convexity of the exponential function, $e^x \geq x+1$ for all $x\in\mathbb{R}$; applying this to the RHS of (6), it follows that

$$\frac{\mathrm{Tr}(W_{t+1})}{\mathrm{Tr}(W_t)} \leq \exp\left(-\beta\mathrm{Tr}\left(L_t\omega_t\right) + \beta^2\mathrm{Tr}\left(L_t^2\omega_t\right)\right). \tag{7}$$

Multiplying the recursive relation (7) for all times $1\leq t\leq T$ and noting that $\mathrm{Tr}(W_1) = \mathrm{Tr}(\mathbb{1}) = d$, we obtain

$$\mathrm{Tr}(W_{T+1}) \leq d\cdot\exp\left(-\beta\sum_{t=1}^{T}\mathrm{Tr}\left(L_t\omega_t\right) + \beta^2\sum_{t=1}^{T}\mathrm{Tr}\left(L_t^2\omega_t\right)\right). \tag{8}$$

Combining the lower bound on $\mathrm{Tr}(W_{T+1})$ in (4) with the upper bound in (8), we have

$$\exp\left(-\beta\lambda_{\min}\left(\sum_{t=1}^{T} L_t\right)\right) \leq d\cdot\exp\left(-\beta\sum_{t=1}^{T}\mathrm{Tr}\left(L_t\omega_t\right) + \beta^2\sum_{t=1}^{T}\mathrm{Tr}\left(L_t^2\omega_t\right)\right),$$

whence taking logarithms yields the relation (3), as desired. $\square$

3 Logarithmic Mistake Bound

To obtain a mistake bound in the quantum online learning setting, it remains to construct loss matrices L_t for each time t such that L_t is Hermitian and satisfies $||L_t||_{\mathrm{Tr}} \leq 1$. For all times t, let us define the convex loss function

$$\ell_t(x) = |x - b_t|, \tag{9}$$

where b_t is the feedback satisfying (1) provided by the environment. Then given learner prediction $\mathrm{Tr}(E_t\omega_t)$, the *loss* of the prediction at time t is given by $\ell_t\left(\mathrm{Tr}(E_t\omega_t)\right)$.

Proposition 3.1. *The loss function $\ell_t(x)$ as defined in (9) satisfies $\ell_t'(x)(x-z) \geq \ell_t(x) - \ell_t(z)$ for any $x, z \in \mathbb{R}$.*

Proof. See Theorem 2.1.3 in [3] for full details. This inequality is due to the global convexity of $\ell_t(x)$, as it states that the slope of a line through two points on the graph of $\ell_t(x)$ is bounded between the slopes of the tangent lines at the endpoints. □

Definition 3.2. For a sequence of learner estimates $\{\omega_t\}_{t\leq T}$ and environment-provided binary measurement operators $\{E_t\}_{t\leq T}$, the *regret* over $T > 0$ rounds of learning is given by

$$R(T) = \sum_{t=1}^{T} \ell_t\left(\mathrm{Tr}(E_t\omega_t)\right) - \min_{\varphi\in\mathbb{D}(d)} \sum_{t=1}^{T} \ell_t\left(\mathrm{Tr}(E_t\varphi)\right).$$

Lemma 3.3. *Let $\{\omega_t\}_{t\geq 1}$ be the estimates of ρ chosen according to the MMW algorithm with loss function $L_t = \ell_t'(\mathrm{Tr}(E_t\omega_t))E_t$ for all times t. Then the regret over T rounds satisfies*

$$R(T) \leq 4\frac{\log(d)}{\varepsilon} + T\frac{\varepsilon}{4}. \tag{10}$$

Proof. First, note that from (9), we have $\ell_t'(x) = \frac{x-b_t}{|x-b_t|} \implies |\ell_t'(x)| = 1$. As a result, $||L_t||_{\mathrm{Tr}} = ||E_t||_{\mathrm{Tr}} \leq 1$ using Definition 1.10; moreover, E_t Hermitian implies L_t is Hermitian, so the defined loss matrices can be applied in the context of the MMW algorithm.

Now, given (3) holds, we can rewrite its RHS in a manner that evokes the regret formula. Notice that for any Hermitian matrix M, $\lambda_{\min}(M) = \min_{\varphi\in\mathbb{D}(d)} \mathrm{Tr}(M\varphi)$, with equality holding when φ is the outer product of an eigenvector corresponding to $\lambda_{\min}(M)$ with itself. Then $\lambda_{\min}\left(\sum_{t=1}^{T} L_t\right) = \min_{\varphi\in\mathbb{D}(d)} \sum_{t=1}^{T} \mathrm{Tr}(L_t\varphi)$. Moreover, since $||L_t||_{\mathrm{Tr}} \leq 1 \implies L_t \preceq \mathbb{1}$ by Lemma 1.6, using Corollary 1.4 repeatedly we have $\mathrm{Tr}\left(L_t^2\omega_t\right) \leq \mathrm{Tr}\left(L_t\omega_t\right) \leq \mathrm{Tr}\left(\omega_t\right) = 1$; hence, $\sum_{t=1}^{T} \mathrm{Tr}\left(L_t^2\omega_t\right) \leq T$.

Inputting these bounding simplifications into (3), we have

$$\sum_{t=1}^{T} \mathrm{Tr}(L_t\omega_t) - \min_{\varphi\in\mathbb{D}(d)} \sum_{t=1}^{T} \mathrm{Tr}(L_t\varphi) \leq \beta T + \frac{\log(d)}{\beta}$$

whence

$$\sum_{t=1}^{T} \ell_t'(\mathrm{Tr}(E_t\omega_t))\mathrm{Tr}(E_t\omega_t) - \min_{\varphi\in\mathbb{D}(d)} \sum_{t=1}^{T} \ell_t'(\mathrm{Tr}(E_t\omega_t))\mathrm{Tr}(E_t\varphi) \leq \beta T + \frac{\log(d)}{\beta}.$$

Lemma 3.1 gives the lower bound

$$\ell_t'(\mathrm{Tr}(E_t\omega_t))\left(\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\varphi)\right) \geq \ell_t(\mathrm{Tr}(E_t\omega_t)) - \ell_t(\mathrm{Tr}(E_t\varphi)),$$

so it follows that

$$R(T) = \sum_{t=1}^{T} \ell_t \left(\mathrm{Tr}(E_t \omega_t)\right) - \min_{\varphi \in \mathbb{D}(d)} \sum_{t=1}^{T} \ell_t \left(\mathrm{Tr}(E_t \varphi)\right) \leq \beta T + \frac{\log(d)}{\beta}.$$

Setting $\beta = \frac{\varepsilon}{4}$ gives the desired inequality (10). ☐

With this lemma in place, we can finally prove the main theorem of the paper.

Proof of Theorem 1.11. We consider a slightly more sophisticated modification of the MMW algorithm strategy for generating estimates ω_t for ρ, inspired by *conservative mistake-bounded algorithms* [9], which only update predictions when a mistake occurs. Set $\mathcal{T} = \{t | \ell_t(\mathrm{Tr}(E_t \omega_t)) > \frac{2\varepsilon}{3}\}$, where $\mathcal{T}$ is updated with a new time sequentially whenever the loss exceeds $\frac{2\varepsilon}{3}$. For $t \notin \mathcal{T}$, since $|b_t - \mathrm{Tr}(E_t \rho)| \leq \frac{\varepsilon}{3}$, the triangle inequality implies $|\mathrm{Tr}(E_t \omega_t) - \mathrm{Tr}(E_t \rho)| \leq \varepsilon$; in other words, for $t \notin \mathcal{T}$, we are guaranteed to not have made a mistake.

Hence, our sequence of estimates $\{\omega_t\}_{t \geq 1}$ is such that when $t \notin \mathcal{T}$, the running weight matrix satisfies $W_{t+1} = W_t$, so $\omega_{t+1} = \omega_t$. When $t \in \mathcal{T}$, we update according to (2) but restricted only to $\mathcal{T}$:

$$W_{t+1} = \exp\left(-\beta \sum_{i \in \mathcal{T}, i \leq t} L_i\right), \qquad \omega_{t+1} = \frac{W_{t+1}}{\mathrm{Tr}(W_{t+1})}.$$

Indeed, since the loss is received after the estimate-based prediction is revealed to the environment, if the loss exceeds the $\frac{2\varepsilon}{3}$ threshold the next estimate is updated. This modification is just a reindexing over $\mathcal{T}$, so our previous results on the MMW algorithm still apply. Now, let us run this learning procedure until some time T, and let $T' = |\{1, \cdots, T\} \cap \mathcal{T}|$ be the number of times an update is made. Clearly $\sum_{t \in \mathcal{T}, t \leq T} \ell_t(\mathrm{Tr}(E_t \omega_t)) \geq \frac{2T'\varepsilon}{3}$ by definition of $\mathcal{T}$. At the same time, due to the feedback guarantee (1), $\sum_{t \in \mathcal{T}, t \leq T} \ell_t(\mathrm{Tr}(E_t \rho)) \leq \frac{T'\varepsilon}{3}$, so this sustains an upper bound for the sum of losses for predictions on a fixed state: $\min_{\varphi \in \mathbb{D}(d)} \sum_{t \in \mathcal{T}, t \leq T} \ell_t \left(\mathrm{Tr}(E_t \varphi)\right) \leq \frac{T'\varepsilon}{3}$. Hence,

$$R(T') = \sum_{t \in \mathcal{T}, t \leq T} \ell_t \left(\mathrm{Tr}(E_t \omega_t)\right) - \min_{\varphi \in \mathbb{D}(d)} \sum_{t \in \mathcal{T}, t \leq T} \ell_t \left(\mathrm{Tr}(E_t \varphi)\right) \geq \frac{T'\varepsilon}{3}.$$

However, from Lemma 3.3, we have

$$4 \frac{\log(d)}{\varepsilon} + T' \frac{\varepsilon}{4} \geq R(T') \geq T' \frac{\varepsilon}{3} \implies T' \leq 48 \frac{\log(d)}{\varepsilon^2}.$$

In particular, taking $T \to \infty$, we find that $|\mathcal{T}| = O\left(\frac{\log(d)}{\varepsilon^2}\right)$. Since mistakes only occur at times in $\mathcal{T}$, it follows that the number of mistakes committed by the learner is also $O\left(\frac{\log(d)}{\varepsilon^2}\right)$, as desired. ☐

4 Future Directions

We have constructed a mistake-bounded algorithm for online quantum state learning that makes at most $O\left(\frac{\log(d)}{\varepsilon^2}\right)$ mistakes in predicting expectation values of binary measurements. The success of this scheme relies on the feedback b_t, which is guaranteed to be within $\frac{\varepsilon}{3}$ of the true expectation value on the unknown state ρ. However, in a practical setting, the existence of such reliable feedback

b_t could be difficult or expensive to construct. As mentioned earlier, the most direct method approximating $\mathrm{Tr}(E\rho)$ for some binary measurement E takes the sample average of positive results of E on several copies of ρ. The issue here is that the sample complexity scales linearly with the number of rounds of online learning. We are guaranteed after $O\left(\frac{\log(d)}{\varepsilon^2}\right)$ mistakes that our estimate of any measurement expectation on ρ is within ε distance from the true values; however, we are not guaranteed how quickly all these mistakes will appear during the online learning process. We could have arbitrarily long sequences of binary measurements that do not require updating our estimate ω_t, but in the process, we are "wasting" copies of ρ on generating feedback for rounds where mistakes do not occur.

Hence, suppose now that the learner has the option to choose the sequence of binary measurements $E_1, E_2, \cdots$ over time, and maintains a time series of estimates $\{\omega_t\}_{t\geq 1}$ of the hidden state ρ as well. The learner can query the environment for feedback b_t that approximates $\mathrm{Tr}(E_t\rho)$ up to error ε; assume this demands some finite sample complexity per round. Let $T_\rho(\delta)$ denote the earliest time such that with probability at least $1-\delta$, we have $|\mathrm{Tr}(E\omega_{T_\rho(\delta)}) - \mathrm{Tr}(E\rho)| \leq \varepsilon$ for all binary measurements E.

Question: (Original) Is there some adaptive scheme for choosing $E_1, E_2, \cdots$ such that $T_\rho(\delta) = p\left(\log(d), \frac{1}{\varepsilon}, \frac{1}{\delta}\right)$, for some polynomial p?

The answer to such a question unites quantum online learning with the notion of sample complexity, potentially offering a sample-efficient method to learning robust estimates of hidden quantum states.

5 Acknowledgements

I'd like to thank Professor Anshu and Chi-Ning Chou for their valuable suggestions regarding the project.

References

[1] S. Aaronson et. al. Online Learning of Quantum States. *Advances in Neural Information Processing Systems 31*, 2018. `arXiv:1802.09025`.

[2] S. Aaronson. The learnability of quantum states. *Proc. Roy. Soc. London*, A463(2088): 3089–3114, 2007. arXiv:quant-ph/0608142.

[3] A. Lowe. Learning Quantum States Without Entangled Measurements. *UWSpace*, 2021. http://hdl.handle.net/10012/17663.

[4] Y. Freund and R. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*(29): 79–103, 1999.

[5] Y. Freund and R. Schapire. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*(55): 771–780, 1999.

[6] S. Arora and S. Kale. A combinatorial, primal-dual approach to semidefinite programs. *Proceedings of the 39th annual ACM symposium on Theory of computing*, 2007.

[7] D. Petz. A survey of trace inequalities. *Functional Analysis and Operator Theory*, 287–298, 1994.

[8] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2010.

[9] N. Littlestone. Learning Quickly When Irrelevant Attributes Abound: A New Linear-threshold Algorithm. *Machine Learning*(2): 285–318, 1988.

6 Appendix

Lemma 6.1. *For all real x such that $|x| \leq 1$, the inequality* $\exp(-x) \leq 1 - x + x^2$ *holds.*

Proof. Let $f(x) = 1 - x + x^2 - \exp(-x)$. Notice that $f'(x) = -1 + 2x + \exp(-x)$, with $f(-1) = 3 - e > 0$. We want to show that $f'(x) < 0$ for $-1 \leq x < 0$ and $f'(x) > 0$ for $x \geq 0$. Since $f(0) = 0$, this will show that $f(x) \geq 0$ for $|x| \leq 1$.

Let us look at $f''(x) = 2 - \exp(-x)$. We have that $f''(x) < 0$ for $x < -\log 2$, and $f''(x) > 0$ otherwise. Hence, $f'(x)$ is decreasing for $x < -\log 2$ and increasing afterwards. Since $f'(-\log 2) = 1 - \log 4 < 0$ but $\lim_{x \to \pm\infty} f'(x) = \infty$, it follows that $f'(x)$ intersects the x-axis at 2 points, one of them being 0. Since $f'(-1) = e - 3 < 0$, by the intermediate value theorem, $f'(x)$ intersects the x-axis at $x < -1$, so we have that $f'(x) < 0$ for $-1 \leq x < 0$. Thus $f(x)$ is decreasing from $x = -1$ to $x = 0$, but $f(0) = 0$ implies that $f(x) \geq 0$ for $-1 \leq x \leq 0$. Now, beyond 0, since $f'(x)$ is increasing for $x > -\log 2$ and $f'(0) = 0$, it follows that $f'(x) > 0$ for $x > 0$. Hence $f(x)$ is increasing for $x > 0$, and since $f(0) = 0$, this implies $f(x) \geq 0$ for $x > 0$.

We have thus shown that $f(x) \geq 0$ for all $x \geq -1$, and the desired inequality follows. $\square$