Online Learning of Quantum States

Scott Aaronson [*]
UT Austin
aaronson@cs.utexas.edu

Xinyi Chen
Princeton University
xinyic2015@gmail.com

Elad Hazan
Princeton University and Google Brain
ehazan@cs.princeton.edu

Ashwin Nayak [†]
University of Waterloo
ashwin.nayak@uwaterloo.ca

Abstract

Suppose we have many copies of an unknown n-qubit state ρ. We measure some copies of ρ using a known two-outcome measurement E_1, then other copies using a measurement E_2, and so on. At each stage t, we generate a current hypothesis σ_t about the state ρ, using the outcomes of the previous measurements. We show that it is possible to do this in a way that guarantees that $|\mathrm{Tr}(E_i\sigma_t) - \mathrm{Tr}(E_i\rho)|$, the error in our prediction for the next measurement, is at least ε at most $\mathrm{O}\big(n/\varepsilon^2\big)$ times. Even in the "non-realizable" setting—where there could be arbitrary noise in the measurement outcomes—we show how to output hypothesis states that do significantly worse than the best possible states at most $\mathrm{O}\Big(\sqrt{Tn}\Big)$ times on the first T measurements. These results generalize a 2007 theorem by Aaronson on the PAC-learnability of quantum states, to the online and regret-minimization settings. We give three different ways to prove our results—using convex optimization, quantum postselection, and sequential fat-shattering dimension—which have different advantages in terms of parameters and portability.

1 Introduction

How many measurements are needed to "learn" an unknown n-qubit quantum state ρ? If we want to reconstruct the full $2^n \times 2^n$ density matrix, even approximately, and if we make no assumptions about ρ, then it is straightforward to show that the number of measurements needed grows exponentially with n. Suppose, on the other hand, that there is some probability distribution $\mathcal{D}$ over possible yes/no measurements, where we identify the measurements with $2^n \times 2^n$ Hermitian matrices E with eigenvalues in $[0, 1]$. Further suppose we are only concerned about learning the state ρ well enough to predict the outcomes of *most* measurements E drawn from $\mathcal{D}$—where "predict" means approximately calculating the probability, $\mathrm{Tr}(E\rho)$, of a "yes" result. Then for how many (known) sample measurements E_i, drawn independently from $\mathcal{D}$, do we need to know the approximate value of $\mathrm{Tr}(E_i\rho)$, before we have enough data to achieve this?

Aaronson [3] proved that the number of sample measurements needed, m, grows only *linearly* with the number of qubits n. What makes this surprising is that it represents an exponential reduction compared to full quantum state tomography. Furthermore, the prediction strategy is extremely simple. Informally, we merely need to find any "hypothesis state" σ that satisfies $\mathrm{Tr}(E_i\sigma) \approx \mathrm{Tr}(E_i\rho)$ for all the sample measurements $E_1, \ldots, E_m$. Then with high probability over the choice of sample measurements, that hypothesis σ will necessarily "generalize," in the sense that $\mathrm{Tr}(E\sigma) \approx \mathrm{Tr}(E\rho)$ for most additional E's drawn from $\mathcal{D}$. The learning theorem led to followup work including a full characterization of quantum advice [6]; efficient learning for stabilizer states [14]; the "shadow tomography" protocol [5]; and recently, the first experimental demonstration of quantum state PAC-learning [15].

[*] Supported by a Vannevar Bush Faculty Fellowship from the US Department of Defense. Part of this work was done while the author was supported by an NSF Alan T. Waterman Award.

[†] Research supported in part by NSERC Canada.

A major drawback of the learning theorem due to Aaronson is the assumption that the sample measurements are drawn *independently* from $\mathcal{D}$—and moreover, that the same distribution $\mathcal{D}$ governs both the training samples, and the measurements on which the learner's performance is later tested. It has long been understood, in computational learning theory, that these assumptions are often unrealistic: they fail to account for adversarial environments, or environments that change over time.

So it is desirable to give learning algorithms that work in the more stringent *online learning* model. Here the learner is presented a sequence of input points, say $x_1, x_2, \ldots$, one at a time. Crucially, there is no assumption whatsoever about the x_t's: the sequence could be chosen adversarially, and even adaptively, which means that the choice of x_t might depend on the learner's behavior on $x_1, \ldots, x_{t-1}$. The learner is trying to learn some unknown function $f(x)$, about which it initially knows only that f belongs to some hypothesis class $\mathcal{H}$—or perhaps not even that; we will also consider the scenario where the learner is just trying to compete with the best predictor in $\mathcal{H}$, which might or might not be a good predictor. The learning proceeds as follows: for each t, the learner first guesses a value y_t for $f(x_t)$, and is then told the true value of $f(x_t)$. Our goal is to design a learning algorithm with the following guarantee: *regardless of the sequence of x_t's, the learner's guess, y_t, will be far from the true value $f(x_t)$ at most k times* (where k, of course, is as small as possible). The x_t's on which the learner errs could be spaced arbitrarily; all we require is that they be bounded in number.

This leads to the following question: can the learning theorem established by Aaronson [3] be generalized to the online learning setting? In other words: is it true that, given a sequence $E_1, E_2, \ldots$ of yes/no measurements, where each E_t is followed shortly afterward by the value of $\mathrm{Tr}(E_t\rho)$, there is a way to anticipate the $\mathrm{Tr}(E_t\rho)$ values by guesses $y_t \in [0,1]$, in such a way that $|y_t - \mathrm{Tr}(E_t\rho)| > \varepsilon$ at most, say, $\mathrm{O}(n)$ times (where $\varepsilon > 0$ is some constant, and n again is the number of qubits)?

The purpose of this paper is to provide an affirmative answer. Throughout the paper, we specify a (two-outcome) measurement of an n qubit mixed state ρ by a "POVM element": that is, a $2^n \times 2^n$ Hermitian matrix E with eigenvalues in $[0,1]$, which "accepts" ρ with probability $\mathrm{Tr}(E\rho)$ and "rejects" ρ with probability $1 - \mathrm{Tr}(E\rho)$.

Theorem 1 *Let ρ be an n-qubit mixed state, and let $E_1, E_2, \ldots$ be a sequence of 2-outcome measurements that are revealed to the learner one by one, each followed by a value $b_t \in [0,1]$ such that $|\mathrm{Tr}(E_t\rho) - b_t| < \varepsilon/2$. Then there is an explicit strategy for outputting hypothesis states $\sigma_1, \sigma_2, \ldots$ such that*

$$|\mathrm{Tr}(E_t\sigma_t) - \mathrm{Tr}(E_t\rho)| > \varepsilon$$

for at most $\mathrm{O}\left(\frac{n}{\varepsilon^2}\right)$ values of t.

We also prove a theorem for the so-called *regret minimization model* (i.e., the "non-realizable case"), where we make no assumption about the input data arising from an actual quantum state, and our goal is simply to do not much worse than the best hypothesis state that could be found with perfect foresight:

Theorem 2 *Let $E_1, E_2, \ldots$ be a sequence of 2-outcome measurements on an n-qubit state. Suppose the E_t's are presented to a learner one-by-one, and that at each iteration, after seeing E_t, the learner is challenged to output a hypothesis state σ_t, and is then told a value $b_t \in [0,1]$, and suffers a "loss" equal to $\ell_t := (\mathrm{Tr}(E_t\sigma_t) - b_t)^2$. Define the "regret" R_T, after T iterations, to be the amount by which the actual loss exceeds the loss of the best single hypothesis:*

$$R_T := \sum_{t=1}^{T} \ell_t - \min_{\sigma} \sum_{t=1}^{T} (\mathrm{Tr}(E_t\sigma) - b_t)^2.$$

Then there is an explicit learning strategy that guarantees $R_T = \mathrm{O}(\sqrt{Tn})$ for all T. This is so even assuming the E_t's and b_t's are chosen adaptively, in response to the learner's previous behavior.

Also, if we care about L_1 loss rather than L_2 loss, then we can achieve a regret of $\mathrm{O}(\sqrt{Tn}\log^{1/2} T)$, although in that case our learning strategy is not explicit.

It is natural to wonder whether Theorems 1 and 2 leave any room for improvement. Theorem 1 is asymptotically optimal in its mistake bound of $\mathrm{O}(n/\varepsilon^2)$; this follows from the property that n-qubit quantum states, considered as a

hypothesis class, have ε-fat-shattering dimension $\Theta(n/\varepsilon^2)$ (see for example [3]). On the other hand, there is room to improve Theorem 2: the only obvious lower bounds are $\Omega(\sqrt{Tn})$ for the L_1 loss and $\Omega(n)$ for the L_2 loss. (These come from considering quantum mixed states that consist of n independent classical coins, each of which could land heads with probability either $1/2$ or $1/2+\varepsilon$.) Thus, it remains both to achieve the "right" L_2 loss, and also to achieve the right L_1 loss via an explicit algorithm and without the log factors.

Let us mention an application of Theorem 1, to appear in simultaneous work. Aaronson [5] has given an algorithm for the so-called *shadow tomography* problem. Here we have an unknown D-dimensional pure state ρ, as well as known two-outcome measurements $E_1, \ldots, E_m$. Our goal is to approximate $\mathrm{Tr}(E_i\rho)$, for *every* i, to within additive error ε. We want to do this by measuring $\rho^{\otimes k}$, where k is as small as possible. Surprisingly, Aaronson showed that this can be achieved with $k = \widetilde{\mathrm{O}}((\log M)^4(\log D)/\varepsilon^5)$: that is, a number of copies of ρ that's only *polylogarithmic* in both D and M. One component of his algorithm is essentially tantamount to online learning with $\widetilde{\mathrm{O}}(n/\varepsilon^3)$—i.e., what we will do in Section 4 of this paper. However, by using Theorem 1 from this paper in a black-box manner, one can improve the sample complexity of shadow tomography to $\widetilde{\mathrm{O}}((\log M)^4(\log D)/\varepsilon^4)$. Details will appear in [5].

To maximize insight, in this paper we give *three* very different approaches to proving Theorems 1 and 2 (although we will not prove every statement with all three approaches).

Our first approach is to adapt techniques from online convex optimization to the setting of complex-valued density matrices. This requires extending standard techniques to cope with convexity and Taylor approximations, which are widely used for functions over the real domain, but not over the complex domain. We also give an efficient iterative algorithm to produce predictions. This approach connects our problem to the modern mainstream of online learning algorithms, and typically achieves the best parameters.

Our second approach is via a postselection-based learning procedure, which starts with the maximally mixed state as a hypothesis and then repeatedly refines it by simulating postselected measurements. This approach builds on earlier work due to Aaronson [1], specifically his proof of $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$. The advantage is that it is almost entirely self-contained, requiring no "power tools" from convex optimization or learning theory. On the other hand, the approach does not give optimal parameters, and we do not know how to prove Theorem 2 with it.

Our third approach is via an upper-bound on the so-called *sequential fat-shattering dimension* of quantum states, considered as a hypothesis class. In the original quantum PAC-learning theorem by Aaronson, the key step was to upper-bound the so-called *ε-fat-shattering dimension* of quantum states considered as a hypothesis class. Fat-shattering dimension is a real-valued generalization of VC dimension. One can then appeal to known results to get a sample-efficient learning algorithm. For online learning, however, bounding the fat-shattering dimension no longer suffices; one instead needs to consider a possibly-larger quantity called sequential fat-shattering dimension. However, by generalizing a lower bound due to [12, 7] for quantum random access codes, we are able to upper-bound the fat-shattering dimension of quantum states. Using known results—in particular, those due to Rakhlin, Sridharan, and Tewari [13]—this implies Theorem 2. We include this proof because the statement that *the hypothesis class of n-qubit states has sequential fat-shattering dimension* $\mathrm{O}(n)$ might be of independent interest: among other things, it implies that *any* online learning algorithm that works given bounded sequential fat-shattering dimension, will work for online learning of quantum states.

To prove an upper bound on sequential fat-shattering dimension, in turn, requires us to generalize the lower bound due to Nayak for quantum random access codes to objects that we call *measurement decision trees*. Roughly speaking, we need to show that, using an n-qubit quantum state ρ, it is possible to store at most $\mathrm{O}(n)$ classical bits, in such a way that any classical bit of our choice can be retrieved with large bias by measuring ρ—and that moreover, this remains true *even if the measurement basis, by which we attempt to retrieve the i^{th} bit, can depend on the values of the first $i-1$ bits (which we need not know).*

1.1 Structure of the paper

We start by describing background and the technical learning setting as well as notations used throughout. In Section 3 we give the algorithms and main theorems derived using convexity arguments and online convex optimization. In Section 4 we describe the postselection algorithm and state the main theorem using this argument. In Section 5 we give a sequential fat-shattering dimension bound for quantum states and its implication for online learning of quantum states.

2 Preliminaries and definitions

We define the trace norm of a matrix M as $\|M\|_{\mathrm{Tr}} \coloneqq \mathrm{Tr}\sqrt{MM^\dagger}$, where $M^\dagger$ is the adjoint of M. We denote the i'th eigenvalue of a Hermitian matrix X by $\lambda_i(X)$, its minimum eigenvalue by $\lambda_{\min}(X)$, and its maximum eigenvalue by $\lambda_{\max}(X)$.

A learner has access to measurements of a quantum state $\rho \in C_n$, where C_n is the set of all trace-1 Hermitian PSD matrices of dimension 2^n:

$$C_n = \{M \in \mathbb{C}^{2^n \times 2^n} \ , \ M = M^\dagger \ , \ M \succeq 0 \ , \ \mathrm{Tr}(M) = 1\} \ .$$

Note that C_n is a convex set. A measurement of ρ is given by a $2^n \times 2^n$ Hermitian matrix E with eigenvalues in $[0, 1]$, and is presented via a classical description of the matrix. The measurement E accepts ρ with probability $\mathrm{Tr}(E\rho)$, and rejects with probability $1 - \mathrm{Tr}(E\rho)$. Our goal is to learn a hypothesis state $\sigma \in C_n$ such that $\mathrm{Tr}(E\sigma) \approx \mathrm{Tr}(E\rho)$ for most measurements E that will be applied in the future.

Online learning and regret. In online learning of quantum states, the learner is required to iteratively predict a state σ_t. It then suffers a "loss" that depends on a measurement E_t presented by an adversary, and is denoted by $f_t(\sigma_t)$. A commonly used loss function is the mean square error given by

$$f_t(\sigma_{t-1}) \coloneqq (\mathrm{Tr}(E_t\sigma_{t-1}) - \mathrm{Tr}(E_t\rho))^2 \ ,$$

where ρ is a fixed quantum state not known to the learner. However, in general, the adversary's actions need not be consistent with any quantum state. The learner then observes feedback from this measurement, which is also provided by the adversary. The simplest feedback is the realization of a binary random variable Y_t such that

$$Y_t = \begin{cases} 1 & \text{with probability } \mathrm{Tr}(\rho E_t) \quad \text{, and} \\ 0 & \text{with probability } 1 - \mathrm{Tr}(\rho E_t) \ . \end{cases}$$

We would like to design a strategy for updating σ_{t-1} such that over T time steps, the learner's total loss is not much more than that of the hypothetical strategy of outputting the same quantum state ω at every time step, where ω minimizes the total loss with perfect hindsight. Formally this is captured by the notion of *regret*, defined as

$$\text{Regret} \coloneqq \sum_{t=1}^{T} f_t(\sigma_t) - \min_{\omega \in C_n} \sum_{t=1}^{T} f_t(\omega) \ .$$

The sequence of E_t's can be arbitrary, even adversarial, based on the learner's previous actions. Note that if the loss function is given by a fixed state ρ (as in the case of mean square error), the minimum total loss would be 0.

A special case of the online learning setting is called *agnostic learning*; here the measurements E_t are drawn from a fixed and unknown distribution $\mathcal{D}$. The setting is called "agnostic" because we still do not assume that the losses correspond to any actual state ρ.

Online mistake bounds. In some online learning scenarios the quantity of interest is not the mean square error, or some other convex loss, but rather simply the total number of mistakes made. For example, we may be interested in the number of iterations for which the distance of $\mathrm{Tr}(E_t\sigma_t)$ is more than ε-far from $\mathrm{Tr}(E_t\rho)$, where ρ is again a fixed state not known to the learner. More formally, let

$$f_t(\sigma_{t-1}) \coloneqq |\,\mathrm{Tr}(E_t\sigma_{t-1}) - \mathrm{Tr}(E_t\rho)| \ ,$$

Then suppose that the goal is to bound the number of iterations in which $f_t(\sigma_{t-1}) > \varepsilon$, without constraints on the sequence of E_t's. We henceforth assume that in this setting, instead of a measurement outcome Y_t as above, the adversary provides as feedback an approximation $\beta_t \in [0, 1]$ that satisfies $|\,\mathrm{Tr}(E_t\rho) - \beta_t| < \frac{\varepsilon}{3}$.

3 Online learning of quantum states

In this section we first describe the algorithm and main guarantee for the online adversarial case with the mean square error loss, and then give an online mistake bound using very similar techniques.

The algorithm below makes use of von Neumann entropy, which relates to the Matrix Exponentiated Gradient algorithm [16]. We follow the template of the Regularized Follow-the-Leader algorithm (see, for example, [11, Chapter 5]), although it would be interesting to see if the techniques due to Warmuth and Kuzmin [17] extend to the complex domain and Hermitian matrices.

Algorithm 1 Regret minimization for quantum tomography

1: Input: T, $\mathcal{K} \coloneqq C_n$, $\eta < \frac{1}{2}$

2: Set $X_1 \coloneqq 2^{-n}\mathbb{I}$.

3: **for** $t = 1, \ldots, T$ **do**

4: Predict X_t. Consider the following cost function $f_t : \mathcal{K} \to \mathbb{R}$ given by measurement E_t and noisy feedback $Y_t \in [0, 1]$:

$$f_t(X) \coloneqq \left(\mathrm{Tr}(E_t X) - Y_t\right)^2 \ .$$

Define

$$\nabla_t \coloneqq 2\left(\mathrm{Tr}(E_t X_t) - Y_t\right) E_t \ .$$

5: Update decision according to the RFTL rule with von Neumann entropy:

$$X_{t+1} \coloneqq \operatorname*{arg\,min}_{X \in \mathcal{K}} \left\{ \eta \sum_{\tau=1}^{t} \mathrm{Tr}(\nabla_t X) + \sum_{i=1}^{2^n} \lambda_i(X) \log \lambda_i(X) \right\} \ . \tag{1}$$

6: **end for**

Remark: The mathematical program in Eq. (1) is convex, and thus can be solved in polynomial time in the dimension, which is 2^n.

Theorem 3 *The L_2 regret of Algorithm 1 is upper-bounded as*

$$\sum_{t=1}^{T} f_t(X_t) - \min_{\rho \in \mathcal{K}} \sum_{t=1}^{T} f_t(\rho) \leq 4\sqrt{2Tn} \ .$$

Remark 1: If $Y_t \in \{0, 1\}$, we can modify Algorithm 1 to achieve L_1 regret $\mathrm{O}(\sqrt{nT}\,)$ by defining $\nabla_t = \mathrm{sgn}(\mathrm{Tr}(E_t X_t) - Y_t) E_t$. The proof is similar to the one given below.

Remark 2: In the case where $Y_t = 0$ with probability $1 - \mathrm{Tr}(E_t\rho)$ and $Y_t = 1$ with probability $\mathrm{Tr}(E_t\rho)$ for some state ρ, $\mathbb{E}[\nabla_t]$ is the gradient of the loss function where we receive precise feedback $\mathrm{Tr}(E_t\rho)$ instead of Y_t. Hence the expected L_2 regret of Algorithm 1, $\mathbb{E}[\sum_{t=1}^{T}(\mathrm{Tr}(E_t X_t) - \mathrm{Tr}(E_t\rho))^2]$, is bounded by $\mathrm{O}(\sqrt{Tn})$.

The proof of Theorem 3 follows along the lines of [11, Theorem 5.2], except that our domain is complex. Therefore, the mean value theorem does not hold, which means we need to approximate the Bregman divergence instead of replacing it by a norm as in the original proof. Another small subtlety is that convexity needs to be carefully defined with respect to the complex domain.

Proof [Proof of Theorem 3] Define $\nabla f_t(X) \coloneqq 2(\mathrm{Tr}(E_t X) - Y_t)E_t$. We note that $f_t(X)$ is convex in the sense that for all $X \in \mathcal{K}$,

$$f_t(X) - f_t(\rho) \leq \nabla f_t(X) \bullet (X - \rho) \ ,$$

where '$\bullet$' denotes the trace inner-product on $2^n \times 2^n$ complex matrices. To see this, it is simpler to use direct computation since the Taylor expansion of complex-valued functions requires additional care:

$$\begin{aligned} f_t(X) - f_t(\rho) - \nabla f_t(X) \bullet (X - \rho) &= 2 \ \mathrm{Tr}(E_t X) \ \mathrm{Tr}(E_t \rho) - \mathrm{Tr}(E_t X)^2 - \mathrm{Tr}(E_t \rho)^2 \\ &= -(\mathrm{Tr}(E_t X) - \mathrm{Tr}(E_t \rho))^2 \leq 0 \ . \end{aligned}$$

Summing over t,

$$\sum_{t=1}^{T} (f_t(X_t) - f_t(\rho)) \leq \sum_{t=1}^{T} \nabla f_t(X_t) \bullet (X_t - \rho) = \sum_{t=1}^{T} \nabla_t \bullet (X_t - \rho) \ .$$

Define $g_t(X) = \nabla_t \bullet X$, and $g_0(X) = \frac{1}{\eta} R(X)$, where $R(X)$ is the negative von Neumann Entropy of X. Denote $D_R^2 \coloneqq \max_{X,Y \in \mathcal{K}} \{R(X) - R(Y)\}$. By [11, Lemma 5.2], for $U \in \mathcal{K}$, we have

$$\sum_{t=1}^{T} [g_t(X_t) - g_t(U)] \leq \sum_{t=1}^{T} \nabla_t \bullet (X_t - X_{t+1}) + \frac{1}{\eta} D_R^2 \ . \tag{2}$$

Define $\Phi_t(X) = \{\eta \sum_{s=1}^{t} \nabla_s \bullet X + R(X)\}$, then the convex program in line 5 of Algorithm 1 finds the minimizer of $\Phi_t(X)$ in $\mathcal{K}$. The following claim shows that that the minimizer is always positive definite, and it is proven in Appendix B.1.

Claim 4 *For all $t \in \{1, 2, ..., T\}$, we have $X_t \succ 0$.*

For $X \succ 0$, we can write $R(X) = \mathrm{Tr}(X \log X)$, and define

$$\nabla \Phi_t(X) \coloneqq \eta \sum_s \nabla_s + \nabla R(X) = \eta \sum_s \nabla_s + \mathbb{I} + \log X \ .$$

The definition of $\nabla \Phi_t(X)$ is analogous to the gradient of $\Phi_t(X)$ if the function is defined over real symmetric matrices. Moreover, the following condition, similar to the optimality condition over a real domain, is satisfied (for a proof, see Appendix B.2).

Claim 5 *For all $t \in \{1, 2, \ldots, T-1\}$,*

$$\nabla \Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) \geq 0 \ . \tag{3}$$

Denote

$$B_{\Phi_t}(X_t \| X_{t+1}) \coloneqq \Phi_t(X_t) - \Phi_t(X_{t+1}) - \nabla \Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) \ .$$

Then by the Pinsker inequality (see, for example, [10] and the references therein),

$$\frac{1}{2} \|X_t - X_{t+1}\|_{\mathrm{Tr}}^2 \leq \mathrm{Tr}(X_t \log X_t) - \mathrm{Tr}(X_t \log X_{t+1}) = B_{\Phi_t}(X_t \| X_{t+1}) \ .$$

We have

$$\begin{aligned} B_{\Phi_t}(X_t \| X_{t+1}) &= \Phi_t(X_t) - \Phi_t(X_{t+1}) - \nabla \Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) \\ &\leq \Phi_t(X_t) - \Phi_t(X_{t+1}) \\ &= \Phi_{t-1}(X_t) - \Phi_{t-1}(X_{t+1}) + \eta \nabla_t \bullet (X_t - X_{t+1}) \\ &\leq \eta \nabla_t \bullet (X_t - X_{t+1}) \ , \end{aligned} \tag{4}$$

because $\Phi_{t-1}(X_t) \leq \Phi_{t-1}(X_{t+1})$ (X_t minimizes $\Phi_t(X)$). Therefore

$$\frac{1}{2} \|X_t - X_{t+1}\|_{\mathrm{Tr}}^2 \leq \eta \nabla_t \bullet (X_t - X_{t+1}) \ . \tag{5}$$

Let $\|M\|_{\mathrm{Tr}}^*$ denote the dual of the trace norm, i.e., the spectral norm of the matrix M. By Generalized Cauchy-Schwartz,

$$\begin{aligned}\nabla_t \bullet (X_t - X_{t+1}) &\leq \|\nabla_t\|_{\mathrm{Tr}}^* \, \|X_t - X_{t+1}\|_{\mathrm{Tr}} \\ &\leq \|\nabla_t\|_{\mathrm{Tr}}^* \sqrt{2\eta \nabla_t \bullet (X_t - X_{t+1})} \ . \qquad \text{by Eq. (5).}\end{aligned}$$

Rearranging,

$$\nabla_t \bullet (X_t - X_{t+1}) \leq 2\eta \|\nabla_t\|_{\mathrm{Tr}}^{*2} \leq 2\eta G_R^2 \ ,$$

where G_R is an upper bound on $\|\nabla_t\|_{\mathrm{Tr}}^*$. Combining with Eq. (2), we arrive at the following bound

$$\sum_{t=1}^{T} \nabla_t \bullet (X_t - \rho) \leq \sum_{t=1}^{T} \nabla_t \bullet (X_t - X_{t+1}) + \frac{1}{\eta} D_R^2 \leq 2\eta T G_R^2 + \frac{1}{\eta} D_R^2 \ .$$

Taking $\eta = \frac{D_R}{G_R \sqrt{2T}}$, we get $\sum_{t=1}^{T} \nabla_t \bullet (X_t - \rho) \leq 2 D_R G_R \sqrt{2T}$. Going back to the regret bound,

$$\sum_{t=1}^{T} (f_t(X_t) - f_t(\rho)) \leq \sum_{t=1}^{T} \nabla_t \bullet (X_t - \rho) \leq 2 D_R G_R \sqrt{2T} \ .$$

We proceed to show that $D_R = \sqrt{n}$. Let Δ_{2^n} denote the set of probability distributions over $[2^n]$. By definition,

$$D_R^2 = \max_{X,Y \in \mathcal{K}} \{R(X) - R(Y)\} \leq \max_{X \in \mathcal{K}} -R(X) \leq \max_{\lambda \in \Delta_{2^n}} \sum_{i=1}^{2^n} \lambda_i \log \frac{1}{\lambda_i} \leq n \ .$$

Since the dual norm of the trace norm is the spectral norm, we have

$$\|\nabla_t\|_{\mathrm{Tr}}^* = \|2 \left(\mathrm{Tr}(E_t X_t) - Y_t\right) E_t\|_2 \leq 2 \|E_t\|_2 \leq 2 \ .$$

Therefore $\sum_{t=1}^{T} (f_t(X_t) - f_t(\rho)) \leq 4\sqrt{2nT}$. ■

3.1 Mistake bound using online learning

Using online convex optimization with a carefully chosen cost function, we can also give a bound on the overall number of mistakes made. Recall that in this setting, in the t'th iteration, the adversary presents the learner with a measurment operator E_t along with a real number $\beta_t \in [0, 1]$ that approximates $\mathrm{Tr}(E_t \rho)$, where ρ is the unknown quantum state.

Consider the following algorithm.

Algorithm 2 Mistake Minimization for Quantum Tomography

1: Input: T, $\mathcal{K} \coloneqq C_n$, $\eta \coloneqq \frac{\varepsilon}{12}$.
2: Set $X_0 \coloneqq 2^{-n}\mathbb{I}$, $W_0 \coloneqq \mathbb{I}$.
3: **for** $t = 1, \ldots, T$ **do**
4: Predict X_{t-1} and define unobserved loss $f_t : \mathcal{K} \to \mathbb{R}$ given by measurement E_t:

$$f_t(X) \coloneqq \mathbb{1}_{\{|\mathrm{Tr}(E_t X) - \mathrm{Tr}(E_t \rho)| > \varepsilon\}} \left|\mathrm{Tr}(E_t X) - \mathrm{Tr}(E_t \rho)\right| \ .$$

Further define

$$\tilde{\nabla}_t \coloneqq \mathbb{1}_{\left\{|\mathrm{Tr}(E_t X_{t-1}) - \beta_t| > \frac{2\varepsilon}{3}\right\}} \mathrm{sgn}(\mathrm{Tr}(E_t X_{t-1}) - \beta_t) E_t \ .$$

5: Update decision according to Matrix Multiplicative Weights:

$$W_t \coloneqq \exp\left(-\eta \sum_{s=1}^{t} \tilde{\nabla}_s\right) \ , \text{ and}$$

$$X_t \coloneqq \frac{W_t}{\mathrm{Tr}(W_t)} \ .$$

6: **end for**

Theorem 6 *For any $\varepsilon > 0$ and any sequence of measurements (E_t), for each $t \geq 1$, Algorithm 2 outputs a hypothesis state X_t depending on $E_1, \cdots, E_t$, and $\beta_1, \cdots, \beta_t \in [0, 1]$ such that as long as $|\beta_t - \mathrm{Tr}(E_t \rho)| \leq \frac{\varepsilon}{3}$ for every t, we have*

$$|\,\mathrm{Tr}(E_{t+1} X_t) - \mathrm{Tr}(E_{t+1} \rho)| > \varepsilon$$

for at most $\mathrm{O}(\frac{n}{\varepsilon^2})$ values of t.

Proof We show that there exists $c \in \mathbb{R}$ such that for all T,

$$\sum_{t=1}^{T} f_t(X_{t-1}) \leq \frac{cn}{\varepsilon} \ .$$

Define

$$\tilde{f}_t(X) \coloneqq \mathbb{1}_{\left\{|\mathrm{Tr}(E_t X) - \beta_t| > \frac{2\varepsilon}{3}\right\}} \left|\mathrm{Tr}(E_t X) - \mathrm{Tr}(E_t \rho)\right| \ .$$

Since $|\beta_t - \mathrm{Tr}(E_t \rho)| \leq \frac{\varepsilon}{3}$, if $|\mathrm{Tr}(E_t X) - \mathrm{Tr}(E_t \rho)| > \varepsilon$, then $|\,\mathrm{Tr}(E_t X) - \beta_t| > \frac{2\varepsilon}{3}$. Hence it suffices to show that for any T,

$$\frac{1}{2} \sum_{t=1}^{T} \tilde{f}_t(X_{t-1}) \leq \frac{c' n}{\varepsilon}, \tag{6}$$

for some constant c'. Fix $T \geq 1$. The following lemma, proved in Appendix B.4, connects the left hand side of Eq. (6) to the regret bound for the Matrix Multiplicative Weights update method:

Lemma 7 *The total loss given by $\tilde{f}_t$ is bounded as*

$$\frac{1}{2} \sum_{t=1}^{T} \tilde{f}_t(X_{t-1}) \leq \sum_{t=1}^{T} \tilde{\nabla}_t \bullet X_{t-1} - \sum_{t=1}^{T} \tilde{\nabla}_t \bullet \rho - \frac{\varepsilon}{12} \sum_{t=1}^{T} \tilde{\nabla}_t^{\,2} \bullet X_{t-1} \ .$$

Extending the regret bound for Matrix Multiplicative Weights from real symmetric measurement matrices [8, Theorem 5.1] to complex Hermitian measurement matrices, we have

Lemma 8

$$\sum_{t=1}^{T} \tilde{\nabla}_t \bullet X_{t-1} \leq \lambda_{\min}\left(\sum_{t=1}^{T} \tilde{\nabla}_t\right) + \eta \sum_{t=1}^{T} {\tilde{\nabla}_t}^2 \bullet X_{t-1} + \frac{n}{\eta} \ . \tag{7}$$

Using the value of η in the algorithm, and noticing that $\lambda_{\min}(\sum_{t=1}^{T} \tilde{\nabla}_t) \leq \sum_{t=1}^{T} \tilde{\nabla}_t \bullet \rho$, we have

$$\sum_{t=1}^{T} \tilde{\nabla}_t \bullet X_{t-1} \leq \sum_{t=1}^{T} \tilde{\nabla}_t \bullet \rho + \frac{\varepsilon}{12} \sum_{t=1}^{T} {\tilde{\nabla}_t}^2 \bullet X_{t-1} + \frac{12n}{\varepsilon} \ . \tag{8}$$

Therefore

$$\frac{1}{2} \sum_{t=1}^{T} \tilde{f}_t(X_{t-1}) \leq \sum_{t=1}^{T} \tilde{\nabla}_t \bullet X_{t-1} - \sum_{t=1}^{T} \tilde{\nabla}_t \bullet \rho - \frac{\varepsilon}{12} \sum_{t=1}^{T} {\tilde{\nabla}_t}^2 \bullet X_{t-1} \leq \frac{12n}{\varepsilon} \ ,$$

and the total loss given by f_t is bounded as

$$\sum_{t=1}^{T} f_t(X_{t-1}) \leq \sum_{t=1}^{T} \tilde{f}_t(X_{t-1}) \leq \frac{24n}{\varepsilon} \ . \tag{9}$$

So there are at most $\mathrm{O}(\frac{n}{\varepsilon^2})$ iterations in which the loss is greater than ε. ■

4 Learning Using Postselection

In this section, we give a direct route to proving a slightly weaker version of Theorem 1: one that does not need the tools of convex optimization, but only tools intrinsic to quantum information.

We need a slight variant of a well-known result, which Aaronson called the "Quantum Union Bound" (see for example [2, 4]). It may be proven along the lines of [2, Lemma 13]. Given a two-outcome measurement E on n-qubits states, we define an operator $\mathcal{M}$ that post-selects on acceptance by E. Let U be any unitary operation on n qubits and an ancilla qubit in register B that maps states of the form $|\psi\rangle|0\rangle$ to $\sqrt{E}\,|\psi\rangle|0\rangle + \sqrt{\mathbb{I}-E}\,|\psi\rangle|1\rangle$. Let Π_0 be the orthogonal projection $\mathbb{I} \otimes |0\rangle\langle 0|$ on states that equal $|0\rangle$ in register B. Then we define the operator $\mathcal{M}$ as

$$\mathcal{M}(\rho) \coloneqq \frac{1}{\mathrm{Tr}(E\rho)} \ \mathrm{Tr}_B\left(U^{-1}\Pi_0 U\left(\rho \otimes |0\rangle\langle 0|\right) U^{-1}\Pi_0 U\right) \ , \tag{10}$$

if $\mathrm{Tr}(E\rho) \neq 0$, and $\mathcal{M}(\rho) \coloneqq 0$ otherwise. We say that the post-selection succeeds with probability $\mathrm{Tr}(E\rho)$. Note that the operator is trace-preserving on states which are accepted by E with non-zero probability.

Lemma 9 (variant of Quantum Union Bound; [2, 4]) *Suppose we have a sequence of two-outcome measurements $E_1, \ldots, E_k$, such that each E_i accepts a certain mixed state ρ with probability at least $1-\varepsilon$. Consider the corresponding operators $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$ that post-select on acceptance by the respective measurements $E_1, E_2, \ldots, E_k$. Let $\widetilde{\rho}$ denote the state $(\mathcal{M}_k \mathcal{M}_{k-1} \cdots \mathcal{M}_1)(\rho)$ obtained by applying each of the k post-selection operations in succession. Then the probability that all the post-selection operations succeed, i.e., the k measurements all accept ρ, is at least $1 - k\sqrt{2\varepsilon}$. Moreover, $\|\widetilde{\rho} - \rho\|_{\mathrm{Tr}} \leq 2k\sqrt{2\varepsilon}$.*

We note that Wilde [18] has proven a tighter version of Lemma 9, involving the bound $\mathrm{O}(\sqrt{k\varepsilon})$ rather than $\mathrm{O}(k\sqrt{\varepsilon})$. However, we do not need the tighter version here.

We now prove the main result of this section.

Theorem 10 *Let ρ be an unknown n-qubit mixed state, let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements, and let $\varepsilon > 0$. There exists a strategy for outputting hypothesis states $\sigma_0, \sigma_1, \ldots$, where σ_t depends only on $E_1, \ldots, E_t$ and real numbers $\beta_1, \ldots, \beta_t$ in $[0,1]$, such that as long as $|\beta_t - \mathrm{Tr}(E_t\rho)| \leq \varepsilon/4$ for every t, we have*

$$|\mathrm{Tr}(E_{t+1}\sigma_t) - \mathrm{Tr}(E_{t+1}\rho)| > \varepsilon$$

for at most $\mathrm{O}\left(\frac{n}{\varepsilon^3} \log \frac{n}{\varepsilon}\right)$ values of t. Here the E_t's and β_t's can otherwise be chosen adversarially.

Proof Let $\rho^* := \rho^{\otimes k}$ be an amplified version of ρ, which lives in a Hilbert space of dimension $D := 2^{kn}$, for some k to be set later. Throughout, we maintain a classical description of a D-dimensional "amplified hypothesis state" σ_t^*, which is always symmetric under permuting the k registers. Given σ_t^*, our actual n-qubit hypothesis state σ_t is then obtained by simply tracing out $k-1$ of the registers.

Given an amplified hypothesis state σ^*, let E_t^* be a two-outcome measurement that acts on σ^* as follows: it applies the measurement E_t to each of the k registers separately, and accepts if and only if the fraction of measurements that accept equals β_t, up to an additive error at most $\varepsilon/2$.

Here is the learning strategy. Our initial hypothesis, $\sigma_0^* := \mathbb{I}/D$, is the D-dimensional maximally mixed state, corresponding to $\sigma_0 := \mathbb{I}/2^n$. For each $t \geq 1$, we are given descriptions of the measurements $E_1, \ldots, E_t$, as well as real numbers $\beta_1, \ldots, \beta_t$ in $[0,1]$, such that $|\beta_i - \operatorname{Tr}(E_i\rho)| \leq \varepsilon/4$ for all $i \in [t]$. We would like to update our old hypothesis σ_{t-1}^* to a new hypothesis σ_t^*, ideally such that the difference $|\operatorname{Tr}(E_{t+1}\sigma_t) - \operatorname{Tr}(E_{t+1}\rho)|$ is small. We do so as follows:

- Given β_t, as well classical descriptions of σ_{t-1}^* and E_t, decide whether $\operatorname{Tr}\left(E_t^*\sigma_{t-1}^*\right) \geq 1 - \frac{\varepsilon}{4}$.
- If yes, then set $\sigma_t^* := \sigma_{t-1}^*$ (i.e., we do not change the hypothesis).
- Otherwise, let σ_t^* be the state obtained by applying E_t^* to σ_{t-1}^* and postselecting on E_t^* accepting. In other words, $\sigma_t^* := \mathcal{M}(\sigma_{t-1}^*)$, where $\mathcal{M}$ is the operator that post-selects on acceptance by E_t^* (as defined above).

We now analyze this strategy. Call t "good" if $\operatorname{Tr}\left(E_t^*\sigma_{t-1}^*\right) \geq 1 - \frac{\varepsilon}{4}$, and "bad" otherwise. Below, we show that

(i) there are at most $\mathrm{O}\left(\frac{n}{\varepsilon^3}\log\frac{n}{\varepsilon}\right)$ bad t's, and

(ii) for each good t, we have $|\operatorname{Tr}(E_t\sigma_{t-1}) - \operatorname{Tr}(E_t\rho)| \leq \varepsilon$.

We start with claim (i). Suppose there have been ℓ bad t's, call them $t(1), \ldots, t(\ell)$, where $\ell \leq (n/\varepsilon)^5$ (we justify this last assumption later, with room to spare). Then there were ℓ events where we postselected on E_t^* accepting σ_{t-1}^*. We conduct a thought experiment, in which the learning strategy maintains a quantum register initially in the maximally mixed state $\mathbb{I}/D$, and applies the post-selection operator corresponding to E_t to the quantum register whenever t is bad. Let p be the probability that all ℓ of these postselection events succeed. Then by definition,

$$p = \operatorname{Tr}\left(E_{t(1)}^*\sigma_{t(1)-1}^*\right)\cdots\operatorname{Tr}\left(E_{t(\ell)}^*\sigma_{t(\ell)-1}^*\right) \leq \left(1 - \frac{\varepsilon}{4}\right)^{\ell}.$$

On the other hand, suppose counterfactually that we had started with the "true" hypothesis, $\sigma_0^* := \rho^* = \rho^{\otimes k}$. In that case, we would have

$$\begin{aligned}\operatorname{Tr}\left(E_{t(i)}^*\rho^*\right) &= \Pr\left[E_{t(i)} \text{ accepts } \rho \text{ between } \left(\beta_{t(i)} - \frac{\varepsilon}{2}\right)k \text{ and } \left(\beta_{t(i)} + \frac{\varepsilon}{2}\right)k \text{ times}\right] \\ &\geq 1 - 2\,\mathrm{e}^{-2k(\varepsilon/4)^2}\end{aligned}$$

for all i. Here the second line follows from the condition that $\left|\operatorname{Tr}\left(E_{t(i)}\rho\right) - \beta_{t(i)}\right| \leq \varepsilon/4$, together with the Hoeffding bound.

We now make the choice $k := \frac{C}{\varepsilon^2}\log\frac{n}{\varepsilon}$, for some constant C large enough that

$$\operatorname{Tr}\left(E_{t(i)}^*\rho^*\right) \geq 1 - \frac{\varepsilon^{10}}{200n^{10}}$$

for all i. So by Lemma 9, all ℓ postselection events would succeed with probability at least

$$1 - \ell\sqrt{\frac{\varepsilon^{10}}{100n^{10}}} \geq 0.9\ .$$

We may write the maximally mixed state, $\mathbb{I}/D$, as

$$\frac{1}{D}\rho^* + \left(1 - \frac{1}{D}\right)\xi\ ,$$

for some other mixed state ξ. For this reason, even when we start with initial hypothesis $\sigma_0^* = \mathbb{I}/D$, all ℓ postselection events still succeed with probability

$$p \geq \frac{0.9}{D} \ .$$

Combining our upper and lower bounds on p now yields

$$\frac{0.9}{2^{kn}} \leq \left(1 - \frac{\varepsilon}{4}\right)^{\ell}$$

or

$$\ell = \mathrm{O}\left(\frac{kn}{\varepsilon}\right) = \mathrm{O}\left(\frac{n}{\varepsilon^3} \log \frac{n}{\varepsilon}\right),$$

which incidentally justifies our earlier assumption that $\ell \leq (n/\varepsilon)^5$.

It remains only to prove claim (ii). Suppose that

$$\mathrm{Tr}\left(E_t^* \sigma_{t-1}^*\right) \geq 1 - \frac{\varepsilon}{4} \ . \tag{11}$$

Imagine measuring k quantum registers prepared in the joint state σ_{t-1}^*, by applying E_t to each register. Since the state σ_{t-1}^* is symmetric under permutation of the k registers, we have that $\mathrm{Tr}(E_t \sigma_{t-1})$, the probability that E_t accepts the first register, equals the fraction of the k registers that E_t accepts. The bound in Eq. ((11)) means that, with probability at least $1 - \frac{\varepsilon}{4}$ over the measurement outcomes, the fraction of registers which E_t accepts is within $\pm\varepsilon/2$ of β_t. The k measurement outcomes are not necessarily independent, but the fraction of registers accepted never differs from β_t by more than 1. So by the union bound, we have

$$|\mathrm{Tr}(E_t \sigma_{t-1}) - \beta_t| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \frac{3\varepsilon}{4} \ .$$

Hence by the triangle inequality,

$$|\mathrm{Tr}(E_t \sigma_{t-1}) - \mathrm{Tr}(E_t \rho)| \leq \frac{3\varepsilon}{4} + |\beta_t - \mathrm{Tr}(E_t \rho)| \leq \varepsilon \ .$$

■

5 Learning Using Sequential Fat-Shattering Dimension

In this section, we prove regret bounds using the notion of *sequential fat-shattering dimension*. We begin with a generalization of a bound for "random access coding", also known as the Index function problem [12, 7]. In fact, the underlying idea is closer to the one used to analyse "serial encoding" in the context of quantum finite automata. (The serial encoding problem is also called Augmented Index in the literature on streaming algorithms.)

We view a complete binary tree of depth $d \geq 0$ as consisting of vertices $v \in \{0,1\}^{\leq d}$. The root of the tree is labeled by the empty string ϵ and each internal vertex v of the tree has two children $v0, v1$. We formalize the notion of an adaptive sequence of measurements through a "measurement decision tree". The tree specifies the measurement to be applied next, given a prefix of such measurements along with the corresponding outcomes.

Definition 5.1 *Let k be a positive integer. A* measurement decision tree *of depth k is a complete binary tree of depth k, each internal vertex v of which is labeled by a triple (S, i, E), where $S \in \{1, \ldots, k\}^l$ is a sequence of length $l := |v|$ of distinct indices, $i \in \{1, \ldots, k\}$ is an index that does not occur in S, and E is a measurement operator. The sequences associated with the children $v0, v1$ of v (if defined) are both equal to S, i.*

For a k-bit string y, and sequence $S := (i_1, i_2, \ldots, i_l)$ with $0 \leq l \leq k$ and $i_j \in \{1, 2, \ldots, k\}$, let y_S denote the substring $y_{i_1} y_{i_2} \cdots y_{i_l}$. The following theorem bounds the number of bits we may encode in an n-qubit quantum state when an arbitrary bit out of the n may be recovered well via a two-outcome measurement. The bound holds even when the measurement for recovering y_i may depend adaptively on a sequence S of other bits of y, *and therefore on the measurements used to recover them.*

Theorem 11 *Let k and n be positive integers. For all k-bit strings $y \coloneqq y_1 \cdots y_k$, let ρ_y be an n-qubit mixed state that "encodes" y. Suppose there exists a measurement decision tree T of depth k such that for each internal vertex v of T and all $y \in \{0,1\}^k$ with $y_S = v$, where (S, i, E) is the triple associated with the vertex v, we have*

(i) if $y_i = 0$ then $\mathrm{Tr}(E\rho_y) \leq p_v$, and

(ii) if $y_i = 1$ then $\mathrm{Tr}(E\rho_y) \geq 1 - p_v$,

where $p_v \in [0, 1/2]$ is the error in predicting the bit y_i at vertex v. Then $n \geq (1 - \mathrm{H}(p))k$, where H is the binary entropy function, and $p \coloneqq \frac{1}{k}\sum_{l=1}^{k} \frac{1}{2^l} \sum_{v \in \{0,1\}^l} p_v$ is the average error.

Proof Let Y be a uniformly random k-bit string. We define a random permutation Π of $\{1, \ldots, k\}$ correlated with Y that is given by the sequence of measurements in the root to leaf path corresponding to Y. More formally, let $\Pi(1) := i$, where i is the index associated with the root of the measurement decision tree T. For $l \in \{2, \ldots, k\}$, let $\Pi(l) := j$, where j is the index associated with the vertex $Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l-1)}$ of the tree T. Let Q be a quantum register such that the joint state of YQ is

$$\frac{1}{2^k} \sum_{y \in \{0,1\}^k} |y\rangle\langle y| \otimes \rho_y \ .$$

The quantum mutual information between Y and Q is bounded as $\mathrm{I}(Y : Q) \leq |Q| = n$. Imagine having performed the first $l-1$ measurements given by the tree T on state Q and having obtained the correct outcomes $Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l-1)}$. These outcomes determine the index $\Pi(l)$ of the next bit that may be learned. By the Chain Rule, for any $l \in \{1, \ldots, k-1\}$,

$$\begin{aligned}
&\mathrm{I}\big(Y_{\Pi(l)} \cdots Y_{\Pi(k)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l-1)}\big) \\
&= \mathrm{I}\big(Y_{\Pi(l)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l-1)}\big) + \mathrm{I}\big(Y_{\Pi(l+1)} \cdots Y_{\Pi(k)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l)}\big) \ .
\end{aligned}$$

Let E be the operator associated with the vertex $V := Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l-1)}$. By hypothesis, the measurement E predicts the bit $Y_{\Pi(l)}$ with error at most p_V. Using the Fano Inequality, and averaging over the prefix V, we get

$$\mathrm{I}\big(Y_{\Pi(l)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(l-1)}\big) \geq \mathbb{E}_V(1 - \mathrm{H}(p_V)) \ .$$

Applying this repeatedly for $l \in \{1, \ldots, k-1\}$, we get

$$\begin{aligned}
\mathrm{I}(Y : Q) &= \mathrm{I}\big(Y_{\Pi(1)} : Q\big) + \mathrm{I}\big(Y_{\Pi(2)} : Q \mid Y_{\Pi(1)}\big) + \mathrm{I}\big(Y_{\Pi}(3) : Q | Y_{\Pi(1)}Y_{\Pi(2)}\big) \\
&\quad + \cdots + \mathrm{I}\big(Y_{\Pi(k)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)} \cdots Y_{\Pi(k-1)}\big) \\
&\geq \sum_{l=1}^{k} \frac{1}{2^l} \sum_{v \in \{0,1\}^l} (1 - \mathrm{H}(p_v)) \\
&\geq (1 - \mathrm{H}(p))k \ ,
\end{aligned}$$

by concavity of the binary entropy function, and the definition of p. ∎

We may modify the requirements on the measurement operators associated with the vertices in Theorem 11 as follows, without changing the essence of its conclusion. For each vertex v, we may have a measurement operator E' such that the error in predicting bit y_i is bounded as

(iii) if $y_i = 0$ then $\mathrm{Tr}(E'\rho_y) \leq a_v - \varepsilon$, and

(iv) if $y_i = 1$ then $\mathrm{Tr}(E'\rho_y) \geq a_v + \varepsilon$,

where $\varepsilon \in [0, 1/2]$ and $a_v \in [\varepsilon, 1 - \varepsilon]$ is a "pivot point" associated with the vertex v. This is a consequence of the following observation. Suppose we are given a measurement operator E', pivot point a_v, and parameter ε as above. We define a new measurement operator E to be associated with vertex v as

$$E \quad := \quad \begin{cases} \frac{E'}{2a_v} & \text{if } a_v \geq \frac{1}{2} \ , \text{ and} \\ \frac{(1-2a_v)E'}{2(1-a_v)} & \text{if } a_v < \frac{1}{2} \ . \end{cases}$$

We may verify that the operator E satisfies the requirements of Theorem 11 with $p \coloneqq (1-\varepsilon)/2$. The modification may be interpreted as producing a fixed outcome 0 or 1 with some probability depending on a_v, and applying the given measurement E' with the remaining probability, so as to translate the pivot point a_v to $1/2$. The conclusion that we get, in this case, is $n \geq (1 - \mathrm{H}(1/2 - \varepsilon))k$, or equivalently $k = \mathrm{O}(n/\varepsilon^2)$.

Let S be a set of functions $f : U \to [0,1]$. Then, following Rakhlin et al. [13], let the *ε-sequential fat-shattering dimension* of S, or $\mathrm{sfat}_\varepsilon(S)$, be the largest k for which we can construct a complete binary tree T of depth k, such that

- each internal vertex $v \in T$ has associated with it a point $x_v \in U$ and a real $a_v \in [0,1]$, such that we traverse the left subtree if $f(x_v) \leq a_v - \varepsilon$ or the right subtree if $f(x_v) \geq a_v + \varepsilon$, and
- for each leaf vertex $v \in T$, there exists an $f \in S$ that causes us to reach v if we traverse T from the root.

Our work above has proven the following theorem:

Theorem 12 *Let U be the set of two-outcome measurements E on an n-qubit state, and let S be the set of all functions $f : U \to [0,1]$ that have the form $f(E) \coloneqq \mathrm{Tr}(E\rho)$ for some ρ. Then for all $\varepsilon > 0$, we have $\mathrm{sfat}_\varepsilon(S) = \mathrm{O}(n/\varepsilon^2)$.*

Theorem 12 strengthens an earlier result due to [3], which proved the same upper bound for the "ordinary" (non-sequential) fat-shattering dimension of quantum states considered as a hypothesis class.

Now we may use existing results from the literature, which relate sequential fat-shattering dimension to online learnability. In particular, in the non-realizable case, Rakhlin et al. [13] recently showed the following:

Theorem 13 (part of Proposition 9 in [13]) *Let S be a set of functions $f : U \to [0,1]$. Suppose we are sequentially presented elements $x_1, x_2, \ldots \in U$, with each x_t followed by an observed value $b_t \in [0,1]$. Then there exists a learning strategy that lets us output a sequence of hypotheses $f_1, f_2, \ldots \in S$, such that the L_1-regret is upper-bounded by:*

$$\sum_{t=1}^{T} |f_t(x_t) - b_t| \leq \min_{f \in S} \sum_{t=1}^{T} |f(x_t) - b_t| + 2T \inf_{\alpha} \left\{ 4\alpha + \frac{12}{\sqrt{T}} \int_{\alpha}^{1} \sqrt{\mathrm{sfat}_\beta(S) \log\left(\frac{2\mathrm{e}T}{\beta}\right)} \mathrm{d}\beta \right\}.$$

Combining Theorem 12 with Theorem 13 gives us the following:

Corollary 14 *Suppose we are presented with a sequence of two-outcome measurements $E_1, E_2, \ldots$ of an n-qubit state, with each E_t followed by an observed value $b_t \in [0,1]$. Then there exists a learning strategy that lets us output a sequence of hypothesis states $\sigma_1, \sigma_2, \ldots$, such that the L_1-regret after the first T iterations is upper-bounded by:*

$$\sum_{t=1}^{T} |\mathrm{Tr}(E_t \sigma_t) - b_t| \leq \min_{\sigma} \sum_{t=1}^{T} |\mathrm{Tr}(E_t \sigma) - b_t| + \mathrm{O}\left(\sqrt{Tn} \log^{1/2} T\right).$$

Note that the result due to [13] is non-explicit. In other words, by following this approach, we do not derive any specific online learning algorithm for quantum states that has the advertised $\mathrm{O}\left(\sqrt{nT} \log^{1/2} T\right)$ upper bound on L_1-regret; we only prove nonconstructively that such an algorithm exists. On the other hand, an advantage of this approach is that we get a bound on L_1 regret rather than L_2 regret.

We expect that the approach in this section, based on sequential fat-shattering dimension, could also be used to prove a mistake bound for the realizable case, but we leave that to future work.

6 Open Problems

We conclude with some questions arising from this work. Can we tighten Theorem 2—presumably, to achieve an L_1 loss of $\mathrm{O}(\sqrt{Tn})$ and an L_2 loss of $\mathrm{O}(n)$, both via fully explicit algorithms? It would also be interesting to obtain regret bounds in terms of L, the loss of the best quantum state in hindsight, as opposed to T.

In what cases can one do online learning of quantum states, not only with few samples, but also with a polynomial amount of computation? What is the tight generalization of our results to measurements with d outcomes? Is it the case, in online learning of quantum states, that *any* algorithm works, so long as it produces hypothesis states that are approximately consistent with all the data seen so far? Note that none of our three proof techniques seem to imply this general conclusion.

References

[1] S. Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1:1–28, 2005. Earlier version in CCC'2004. quant-ph/0402095.

[2] S. Aaronson. QMA/qpoly is contained in PSPACE/poly: de-Merlinizing quantum protocols. In *Proc. Conference on Computational Complexity*, pages 261–273, 2006. quant-ph/0510230.

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

[4] S. Aaronson. The complexity of quantum states and transformations: From quantum money to black holes, February 2016. Lecture Notes for the 28th McGill Invitational Workshop on Computational Complexity, Holetown, Barbados. With guest lectures by A. Bouland and L. Schaeffer. www.scottaaronson.com/barbados-2016.pdf.

[5] S. Aaronson. Shadow tomography of quantum states. To appear in Proceedings of STOC'2018. arXiv:1711.01053, 2018.

[6] S. Aaronson and A. Drucker. A full characterization of quantum advice. *SIAM J. Comput.*, 43(3):1131–1183, 2014. Earlier version in STOC'2010. arXiv:1004.0377.

[7] A. Ambainis, A. Nayak, A. Ta-Shma, and U. V. Vazirani. Quantum dense coding and quantum finite automata. *J. of the ACM*, 49:496–511, 2002. Combination of ([12]) and earlier version in STOC'1999, pp. 376-383. quant-ph/9804043.

[8] S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

[9] K. M. R. Audenaert and J. Eisert. Continuity bounds on the quantum relative entropy. *Journal of Mathematical Physics*, 46(10):102104, 2005. arXiv:quant-ph/0503218.

[10] E. A. Carlen and E. H. Lieb. Remainder terms for some quantum entropy inequalities. *Journal of Mathematical Physics*, 55(4), 2014. arXiv:1402.3840.

[11] E. Hazan. *Introduction to Online Convex Optimization*, volume 2 of *Foundations and Trends in Optimization*. 2015.

[12] Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proc. IEEE FOCS*, pages 369–376, 1999. quant-ph/9904093.

[13] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning via sequential complexities. *The Journal of Machine Learning Research*, 16(1):155–186, 2015.

[14] A. Rocchetto. Stabiliser states are efficiently PAC-learnable. arXiv:1705.00345, 2017.

[15] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino. Experimental learning of quantum states. arXiv:1712.00127, 2017.

[16] K. Tsuda, G. Rätsch, and M. K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *Journal of Machine Learning Research*, 6:995–1018, 2005.

[17] Manfred K Warmuth and Dima Kuzmin. Bayesian generalized probability calculus for density matrices. *Machine Learning*, 78(1-2):63, 2010.

[18] M. Wilde. Sequential decoding of a general classical-quantum channel. *Proc. Roy. Soc. London*, A469(2157):20130259, 2013. arXiv:1303.0808.

A Auxiliary Lemmas

The following lemma is from [16], given here for completeness.

Lemma 15 *For Hermitian matrices A, B and Hermitian PSD matrix X, if $A \succeq B$, then* $\mathrm{Tr}(AX) \geq \mathrm{Tr}(BX)$.

Proof Let $C \coloneqq A - B$. By definition, $C \succeq 0$. It suffices to show that $\mathrm{Tr}(CX) \geq 0$. Let $VQV^\dagger$ be the eigendecomposition of X, and let $C = VPV^\dagger$, where $P \coloneqq V^\dagger CV \succeq 0$. Then $\mathrm{Tr}(CX) = \mathrm{Tr}(VPQV^\dagger) = \mathrm{Tr}(PQ) = \sum_{i=1}^n P_{ii}Q_{ii}$. Since $P \succeq 0$ and all the eigenvalues of X are nonnegative, $P_{ii} \geq 0$, $Q_{ii} \geq 0$. Therefore $\mathrm{Tr}(CX) \geq 0$. ■

Lemma 16 *If A, B are Hermitian matrices, then* $\mathrm{Tr}(AB) \in \mathbb{R}$.

Proof The proof is similar to Lemma 15. Let $VQV^\dagger$ be the eigendecomposition of A. Then Q is a real diagonal matrix. We have $B = VPV^\dagger$, where $P \coloneqq V^\dagger BV$. Note that $P^\dagger = V^\dagger B^\dagger V = P$, so P has a real diagonal. Then $\mathrm{Tr}(AB) = \mathrm{Tr}(VQV^\dagger VPV^\dagger) = \mathrm{Tr}(VQPV^\dagger) = \mathrm{Tr}(QP) = \sum_{i=1}^n Q_{ii}P_{ii}$. Since $Q_{ii}, P_{ii} \in \mathbb{R}$ for all i, $\mathrm{Tr}(AB) \in \mathbb{R}$. ■

B Proof of Lemmas

B.1 Proof of Claim 4

Proof Let $P \in \mathcal{K}$ be such that $\lambda_{\min}(P) = 0$. Suppose $P = VQV^\dagger$, where Q is a diagonal matrix with real values on the diagonal. Assume that $Q_{1,1} = \lambda_{\max}(P)$ and $Q_{2^n,2^n} = \lambda_{\min}(P) = 0$. Let $P' = VQ'V^\dagger$ such that $Q'_{1,1} = Q_{1,1} - \varepsilon$, $Q'_{2^n,2^n} = \varepsilon$ for $\varepsilon < \lambda_{\max}(P)$, and $Q'_{ii} = Q_{ii}$ for $i \in \{2, 3, ..., 2^n - 1\}$, so $P' \in \mathcal{K}$. We show that there exists $\varepsilon > 0$ such that $\Phi_t(P') \leq \Phi_t(P)$. Expanding both sides of the inequality, we see that it is equivalent to showing that for some ε,

$$\eta \sum_{s=1}^{t} 2(\mathrm{Tr}(E_s X_s) - Y_s)E_s^\dagger \bullet (P' - P) \leq \lambda_1(P)\log\lambda_1(P) - \lambda_1(P')\log\lambda_1(P') - \varepsilon\log\varepsilon \ .$$

Let $L = \lambda_1(P) = Q_{1,1}$, and $A = \eta\sum_{s=1}^t 2(\mathrm{Tr}(E_sX_s) - Y_s)E_s^\dagger$. The inequality then becomes

$$A \bullet (P' - P) \leq L\log L - (L - \varepsilon)\log(L - \varepsilon) - \varepsilon\log\varepsilon \ .$$

Observe that $0 \leq \mathrm{Tr}(E_sX_s) \leq \|E_s\| \leq 1$, so $\|A\| \leq 2\eta\sum_{s=1}^t \|E_s\| \leq 2\eta t$. So by the Generalized Cauchy-Schwartz inequality,

$$A \bullet (P' - P) \leq 2\eta t \|P' - P\|_{\mathrm{Tr}} = 4\varepsilon\eta t \ .$$

Since η, t, L are finite and $-\log\varepsilon \to \infty$ as $\varepsilon \to 0$, there exists ε small such that $4\eta t \leq \log L - \log\varepsilon$. We have

$$\begin{aligned} 4\eta t\varepsilon &\leq \varepsilon\log L - \varepsilon\log\varepsilon \\ &= L\log L - (L - \varepsilon)\log L - \varepsilon\log\varepsilon \\ &\leq L\log L - (L - \varepsilon)\log(L - \varepsilon) - \varepsilon\log\varepsilon \ . \end{aligned}$$

So there exists $\varepsilon > 0$ such that $\Phi_t(P') \leq \Phi_t(P)$. If P has multiple eigenvalues that are 0, we can repeat the proof and show that there exists a PD matrix P' such that $\Phi_t(P') \leq \Phi_t(P)$. Since X_t is a minimizer of Φ_{t-1} and $X_1 \succ 0$, we conclude that $X_t \succ 0$ for all t. ■

B.2 Proof of Claim 5

Proof Suppose $\nabla\Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) < 0$. Let $a \in (0,1)$ and $\bar{X} = (1-a)X_{t+1} + aX_t$, then $\bar{X}$ is a density matrix and is positive definite. Define $\triangle = \bar{X} - X_{t+1} = a(X_t - X_{t+1})$. We have

$$\begin{aligned}\Phi_t(\bar{X}) - \Phi_t(X_{t+1}) &= a\nabla\Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) + B_{\Phi_t}(\bar{X} \| X_{t+1}) \\ &\leq a\nabla\Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) + \frac{\operatorname{Tr}(\triangle^2)}{\lambda_{\min}(X_{t+1})} \\ &= a\nabla\Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) + \frac{a^2 \operatorname{Tr}((X_t - X_{t+1})^2)}{\lambda_{\min}(X_{t+1})} \ .\end{aligned}$$

The above inequality is due to [9, Theorem 2]. Dividing by a on both sides, we have

$$\frac{\Phi_t(\bar{X}) - \Phi_t(X_{t+1})}{a} \leq \nabla\Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) + \frac{a \operatorname{Tr}((X_t - X_{t+1})^2)}{\lambda_{\min}(X_{t+1})} \ .$$

So we can find a small enough such that the right hand side of the above inequality is negative. However, we would have $\Phi_t(\bar{X}) - \Phi_t(X_{t+1}) < 0$, which is a contradiction. So $\nabla\Phi_t(X_{t+1}) \bullet (X_t - X_{t+1}) \geq 0$. ■

B.3 Proof of Lemma 7

Proof If $|\operatorname{Tr}(E_t X_{t-1}) - \beta_t| \leq \frac{2\varepsilon}{3}$, then $\tilde{\nabla}_t = 0$, $\tilde{f}_t(X_{t-1}) = 0$, so we only need to show that if $|\operatorname{Tr}(E_t X_{t-1}) - \beta_t| > \frac{2\varepsilon}{3}$, then

$$\frac{1}{2}\left|\operatorname{Tr}(E_t X_{t-1}) - \operatorname{Tr}(E_t\rho)\right| \leq \tilde{\nabla}_t \bullet X_{t-1} - \tilde{\nabla}_t \bullet \rho - \frac{\varepsilon}{12}{\tilde{\nabla}_t}^2 \bullet X_{t-1} \ . \tag{12}$$

First note that $E_t E_t \preceq E_t$, and since X_{t-1} is Hermitian PSD, $\operatorname{Tr}(X_{t-1}E_tE_t) \leq \operatorname{Tr}(X_{t-1}E_t)$. Suppose $\operatorname{Tr}(E_t X_{t-1}) > \beta_t + \frac{2\varepsilon}{3}$, we have $\operatorname{Tr}(E_t X_{t-1}) > \operatorname{Tr}(E_t\rho) + \frac{\varepsilon}{3}$, so $\tilde{\nabla}_t = E_t$ and Eq. (12) becomes

$$\frac{1}{2}(\operatorname{Tr}(E_t X_{t-1}) - \operatorname{Tr}(E_t\rho)) \leq E_t \bullet X_{t-1} - E_t \bullet \rho - \frac{\varepsilon}{12} E_t E_t \bullet X_{t-1} \ . \tag{13}$$

Now

$$\begin{aligned}&\frac{1}{2}\left(\operatorname{Tr}(E_t X_{t-1}) - \operatorname{Tr}(E_t\rho)\right) - \operatorname{Tr}(E_t X_{t-1}) + \operatorname{Tr}(E_t\rho) + \frac{\varepsilon}{12}\operatorname{Tr}(E_t E_t X_{t-1}) \\ &\quad\leq \left(\frac{\varepsilon}{12} - \frac{1}{2}\right)\operatorname{Tr}(E_t X_{t-1}) + \frac{1}{2}\operatorname{Tr}(E_t\rho) \\ &\quad\leq \left(\frac{\varepsilon}{12} - \frac{1}{2}\right)\operatorname{Tr}(E_t X_{t-1}) + \frac{1}{2}\left(\operatorname{Tr}(E_t X_{t-1}) - \frac{\varepsilon}{3}\right) \\ &\quad= \frac{\varepsilon}{12}\operatorname{Tr}(E_t X_{t-1}) - \frac{\varepsilon}{6} \\ &\quad\leq \frac{\varepsilon}{12} - \frac{\varepsilon}{6} < 0 \ ,\end{aligned}$$

where second inequality follows from $\operatorname{Tr}(E_t\rho) < \operatorname{Tr}(E_t X_{t-1}) - \frac{\varepsilon}{3}$, and the third from $\operatorname{Tr}(E_t X_{t-1}) \leq 1$. The resulting inequality is equivalent to Eq. (13).

The other case is treated similarly. ■

B.4 Proof of Lemma 8

This matrix multiplicative weights lemma is from [8], and is given here for completeness.

Proof We proceed to bound and bound $\log \mathrm{Tr}(W_T)$ from above and below. First note that for all t, $\tilde{\nabla}_t$ is a Hermitian matrix with spectral norm bounded by 1. Then

$$\begin{aligned}
\mathrm{Tr}(W_T) &= \mathrm{Tr}\left(\exp\left(-\eta \sum_{t=1}^{T} \tilde{\nabla}_t\right)\right) \\
&= \mathrm{Tr}\left(\exp\left(-\eta \sum_{t=1}^{T-1} \tilde{\nabla}_t - \eta \tilde{\nabla}_T\right)\right) \\
&\leq \mathrm{Tr}\Big(W_{T-1} \exp\Big(-\eta \tilde{\nabla}_T\Big)\Big) && \text{by the Golden-Thompson Inequality} \\
&= \mathrm{Tr}(W_{T-1}) X_{T-1} \bullet \exp\Big(-\eta \tilde{\nabla}_T\Big) \quad .
\end{aligned}$$

By induction, we have

$$\log \mathrm{Tr}(W_T) \leq \log \mathrm{Tr}(W_0) + \sum_{t=1}^{T} \log\Big(X_{t-1} \bullet \exp\Big(-\eta \tilde{\nabla}_t\Big)\Big) = n + \sum_{t=1}^{T} \log\Big(X_{t-1} \bullet \exp\Big(-\eta \tilde{\nabla}_t\Big)\Big) \quad .$$

Using the property that $\exp\Big(-\eta \tilde{\nabla}_t\Big) \preceq \mathbb{I} - \eta \tilde{\nabla}_t + \eta^2 \tilde{\nabla}_t^2$, we have

$$\begin{aligned}
\log \mathrm{Tr}(W_T) &\leq n + \sum_{t=1}^{T} \log\Big(X_{t-1} \bullet \Big(\mathbb{I} - \eta \tilde{\nabla}_t + \eta^2 \tilde{\nabla}_t^2\Big)\Big) \\
&= n + \sum_{t=1}^{T} \log\Big(1 - \eta X_{t-1} \bullet \tilde{\nabla}_t + \eta^2 X_{t-1} \bullet \tilde{\nabla}_t^2\Big) && \text{since } \mathrm{Tr}(X_{t-1}) = 1 \\
&\leq n + \sum_{t=1}^{T} \Big(-\eta X_{t-1} \bullet \tilde{\nabla}_t + \eta^2 X_{t-1} \bullet \tilde{\nabla}_t^2\Big) \quad , && \text{since } \log(x) \leq x - 1 \text{ for } x > 0.
\end{aligned}$$

On the other hand,

$$\mathrm{Tr}(W_T) \geq \lambda_{\max}(W_T) = \lambda_{\max}\left(\exp\left(-\eta \sum_{t=1}^{T} \tilde{\nabla}_t\right)\right) = \exp\left(\lambda_{\max}\left(-\eta \sum_{t=1}^{T} \tilde{\nabla}_t\right)\right) \quad .$$

Taking logarithms,

$$\log \mathrm{Tr}(W_T) \geq \lambda_{\max}\left(-\eta \sum_{t=1}^{T} \tilde{\nabla}_t\right) \quad .$$

Combining the upper and lower bounds, we have

$$\lambda_{\max}\left(-\eta \sum_{t=1}^{T} \tilde{\nabla}_t\right) \leq n + \sum_{t=1}^{T} \Big(-\eta X_{t-1} \bullet \tilde{\nabla}_t + \eta^2 X_{t-1} \bullet \tilde{\nabla}_t^2\Big) \quad ,$$

which implies the statement of the lemma. ■