Adaptive Online Learning of Quantum States

Xinyi Chen[1,2], Elad Hazan[1,2], Tongyang Li[3,4], Zhou Lu[1,2], Xinzhao Wang[3,4], and Rui Yang[3,4]

[1]Department of Computer Science, Princeton University, NJ 08540, USA
[2]Google DeepMind Princeton, NJ 08542, USA
[3]Center on Frontiers of Computing Studies, Peking University, 100871 Beijing, China
[4]School of Computer Science, Peking University, 100871 Beijing, China

The problem of efficient quantum state learning, also called shadow tomography, aims to comprehend an unknown d-dimensional quantum state through POVMs. Yet, these states are rarely static; they evolve due to factors such as measurements, environmental noise, or inherent Hamiltonian state transitions. This paper leverages techniques from adaptive online learning to keep pace with such state changes.

The key metrics considered for learning in these mutable environments are enhanced notions of regret, specifically adaptive and dynamic regret. We present adaptive and dynamic regret bounds for online shadow tomography, which are polynomial in the number of qubits and sublinear in the number of measurements. To support our theoretical findings, we include numerical experiments that validate our proposed models.

1 Introduction

As opposed to their classical counterparts, quantum states over n qubits requires 2^n "amplitudes" for their description. Therefore, recovering the density matrix of a quantum system with n qubits from measurements becomes an impractical task for even moderate values of n. The setting of shadow tomography [1] attempts to circumvent this issue: rather than recovering the entire state matrix, the goal is to accurately reconstruct the "shadow" that the density matrix casts on a set of measurements that are known in advance.

More precisely, let ρ be an n-qubit quantum state: that is, a $2^n \times 2^n$ Hermitian positive semidefinite matrix with $\mathrm{Tr}(\rho) = 1$. We can measure ρ by a two-outcome positive operator-valued measure (POVM) E, which can be represented by a $2^n \times 2^n$ Hermitian matrix with eigenvalues in $[0, 1]$. Such a measurement E accepts ρ (returns 1) with probability $\mathrm{Tr}(E\rho)$ and rejects ρ (returns 0) with probability $1 - \mathrm{Tr}(E\rho)$. Given a sequence of POVMs $E_1, \ldots, E_m$, shadow tomography uses copies of ρ to estimate $\mathrm{Tr}(E_i\rho)$ within ϵ for all $i \in [m]$ with probability at least $2/3$.

Xinyi Chen: xinyic@princeton.edu
Elad Hazan: ehazan@princeton.edu
Tongyang Li: tongyangli@pku.edu.cn
Zhou Lu: zhoul@princeton.edu
Xinzhao Wang: wangxz@stu.pku.edu.cn
Rui Yang: ypyangrui@pku.edu.cn

The state-of-the-art result [2] can accomplish this task with $\tilde{O}(n(\log m)^2/\epsilon^4)$ copies of ρ, exponentially improving upon full state tomography in terms of n. However, in the real world a POVM E can be chosen based on previous actions from the observer (learner) and is not known in advance. Moreover, shadow tomography may be unfeasible on near-term quantum computers, as current methods [1, 3, 4, 2] all require a joint measurement on multiple copies of ρ, which needs to be implemented on larger quantum systems limited by the scale of existing quantum devices.

In light of both concerns it is natural to consider *online learning of quantum states*. In this problem, the measurements appear sequentially and the goal is to select a sequence of states to minimize regret compared to the optimal quantum state in retrospect. This problem naturally fits into the *online convex optimization* framework, in which a player makes a decision at each step, and then is revealed a convex loss function and suffers the loss of its decision. The player makes decisions to minimize the total loss. However, this problem is too difficult since the loss functions can be arbitrary, so we only expect to minimize the loss of the player minus the minimum loss of making the same decision at every step, which is called the *regret* of the player. In the online learning of quantum states setting, the player outputs a quantum state φ_t at step t and suffers a loss measuring the difference between φ_t and the true state under the chosen POVM. The sequential nature of the setting gives rise to algorithms that treat each measurement separately, thereby eliminating the need for joint measurements[1]. The existing algorithm of [5], along with subsequent works [6, 7, 8, 9], yields a sequence of states with regret $O(\sqrt{Tn})$ for T iterations. In some cases, these regret bounds can be formalized as *mistake bounds* which count the number of steps that the player suffers a loss larger than a given threshold.

Nevertheless, existing literature on shadow tomography does not cover an important factor in current quantum computers: fluctuation of quantum states. In practical scenarios, we often cannot calibrate the quantum state constantly, and shadow tomography algorithms requiring multiple identical copies of ρ may not be viable. For instance, on the IBM quantum computing platform, a quantum circuit with measurements will be executed hundreds or thousands of shots in a row without calibrations in such consecutive experiments.

In real-world experiments, a quantum state ρ is typically obtained by evolving a control Hamiltonian H for some time τ starting from an initial state ρ_0, i.e., $\rho = e^{iH\tau}\rho_0 e^{-iH\tau}$. If the device is not calibrated for a while, the control Hamiltonian implemented is a noisy one $H + \delta H$, and the actual quantum state we prepare is a perturbed state $\rho' = e^{-i(H+\delta H)\tau}\rho_0 e^{i(H+\delta H)\tau}$. More discussions about noises in quantum states can be found in [10, 11, 12, 13, 14].

In light of these fluctuations, recent study on state tomography extends to temporal quantum tomography [15] with changing quantum states. Therefore, a fundamental question from our perspective of machine learning is learning changing quantum states, which naturally fits into the framework of *adaptive online learning*. Such quantum algorithms are crucial for near-term noisy, intermediate-scale quantum computers (NISQ), since the design of noise-resilient algorithms may extend the computational power of NISQ devices [16].

Adaptive online learning. The motivation above suggests a learning problem that assumes the quantum state prepared at time step $t = 1, 2, \ldots, T$ is ρ_t. Due to imperfect calibration, ρ_t at different time t is different, and our goal is to learn ρ_t in an online fashion. The problem of online learning of a changing concept has received significant attention in

[1] In some shadow tomography algorithms using similar online learning schemes [1, 2], they do not need joint measurements when updating the state but do need them when choosing the POVMs

the machine learning literature, and we adapt several of these techniques to the quantum setting, which involves high-dimensional spaces over the complex numbers.

The first metric we consider is the "dynamic regret" introduced by [17], which measures the difference between the learner's loss and that of a changing comparator. The dynamic regret bounds are usually characterized by how much the optimal comparator changes over time, known as the "path length."

Next, we consider a more sophisticated metric, the "adaptive regret," as introduced by [18]. This metric measures the maximum of the regret over all intervals, essentially taking into account a changing comparator. Many extensions and generalizations of the original technique have been presented in works such as [19, 20, 21]. Minimizing adaptive regret requires a combination of expert algorithms and continuous online learning methods, which we also adopt for the quantum setting.

Contributions. In this paper, we systematically study adaptive online learning of quantum states. In quantum tomography, we might have prior knowledge about ρ_t, but in online learning, it can be adversarial. As such, we expect regret guarantees that extend beyond just comparing with the best fixed $\varphi \in C_n$. We study the following questions summarized in Table 1:

- **Dynamic regret:** We consider minimizing regret under the assumption that the comparator φ changes slowly:

$$\mathrm{Regret}_{T,\mathcal{P}}^{\mathcal{A}} = \sum_{t=1}^{T} \ell_t\left(\mathrm{Tr}(E_t x_t^{\mathcal{A}})\right) - \min_{\varphi_t \in C_n} \sum_{t=1}^{T} \ell_t\left(\mathrm{Tr}(E_t \varphi_t)\right), \text{ where } \mathcal{P} = \sum_{t=1}^{T-1} \|\varphi_t - \varphi_{t+1}\|_1. \quad (1)$$

  Here $\mathcal{P}$ is bounded and defined as the path length, the functions ℓ_t are L-Lipschitz for all $t \in [T]$, and $\|\cdot\|_1$ is the ℓ_1 norm of the singular values of a matrix known as the nuclear norm. We propose an algorithm based on online mirror descent (OMD) that achieves an $\tilde{O}(\sqrt{nT\mathcal{P}})$ regret bound. Although the analysis shares the same spirit as the standard OMD proof, tackling optimization in the complex domain requires new tools from several areas. Specifically, we use Klein's inequality from matrix analysis to study the dual update in OMD, and the Fannes–Audenaert inequality [22] from quantum information theory to obtain the dynamic regret bounds. Compared to [5], our analysis is more general due to the introduction of these tools.

  As an application, under the assumption that φ_t evolves according to a family of dynamical models, we can also prove $\tilde{O}(\sqrt{nT\mathcal{P}'})$ dynamic regret bounds where $\mathcal{P}'$ is a variant of path length defined in Eq. (6). This covers two natural scenarios: either the dynamics comes from a family of M known models (which incurs a $\log M$ overhead), or we know that the dynamical model is an $O(1)$-local quantum channel, a natural assumption for current quantum computers fabricated by 2D chips [23, 24, 25] where a quantum gate can only influence a constant number of qubits.

- **Adaptive regret:** We also consider minimizing the strongly adaptive regret [19] of an algorithm $\mathcal{A}$ at time T for length τ:

$$\text{SA-Regret}_T^{\mathcal{A}}(\tau) = \max_{I \subseteq [T], |I| = \tau} \left(\sum_{t \in I} \ell_t\left(\mathrm{Tr}(E_t x_t^{\mathcal{A}})\right) - \min_{\varphi \in C_n} \sum_{t \in I} \ell_t\left(\mathrm{Tr}(E_t \varphi)\right) \right) \quad (2)$$

  where $x_t^{\mathcal{A}}$ is the output of algorithm $\mathcal{A}$ at time t, and ℓ_t is a real loss function that is revealed to the learner, for example, $\ell_t(z) = |z - b_t|$ and $\ell_t(z) = (z - b_t)^2$ for some

$b_t \in [0,1]$.[2] We show that combining algorithms from [5], [20] gives an $O(\sqrt{n\tau \log(T)})$ bound for adaptive regret. We then derive an $O(\sqrt{knT \log(T)})$ regret bound when the comparator φ is allowed to shift k times.

- **Mistake bound:** Given a sequence of two-outcome measurements $E_1, E_2, \ldots, E_T$ where each E_t is followed afterward by a value b_t such that $|\operatorname{Tr}(E_t\rho_t) - b_t| \leq \epsilon/3$. Assuming the ground truth ρ_t has k-shift or bounded path length $\mathcal{P}$, we build upon the regret bounds to derive upper bounds of the number of mistakes made by an algorithm whose output is a sequence of $x_t^{\mathcal{A}} \in [0,1]$, where $|x_t^{\mathcal{A}} - \operatorname{Tr}(E_t\rho_t)| > \epsilon$ is considered a mistake. For dynamic regret and adaptive regret settings, we give mistake bounds in Corollary 1 and Corollary 4, respectively.

Regret	Reference	Setting	O-Bound
Dynamic	Theorem 1	Path length	$L\sqrt{T(n+\log(T))\mathcal{P}}$
Dynamic	Corollary 2	Family of M dynamical models	$L\sqrt{T(n+\log(T))\mathcal{P}'}+\sqrt{T\log(M\log(T))}$
Dynamic	Corollary 3	l-local quantum channels	$L\sqrt{T(n+\log(T))\mathcal{P}'}+2^{6l}\sqrt{T}$
Adaptive	Lemma 1	Arbitrary interval with length τ	$L\sqrt{n\tau\log(T)}$
Adaptive	Theorem 2	k-shift	$L\sqrt{knT\log(T)}$

Table 1: Adaptive and dynamic regret bounds of online learning of quantum states. Here n is the number of qubits, L is the Lipschitzness of loss functions, T is the total number of iterations, $\mathcal{P}$ is the path length defined in Eq. (1), $\mathcal{P}'$ in Corollary 2 is a variant of path length defined in Eq. (6) and $\mathcal{P}'$ in Corollary 3 is defined in Eq. (15). l-local quantum channels are explained in Section 3.

We also note that our results can be extended from two-outcome POVMs to K-outcome POVMs for $K \geq 2$, inspired by a recent work by [26]. A K-outcome measurements $\mathcal{M}$ can be described by K $2^n \times 2^n$ Hermitian matrices $E_1, \ldots, E_K$ such that $0 \preceq E_j \preceq I$, $\sum_{j=1}^K E_j = I$, and for any quantum state ρ, the measurement $\mathcal{M}$ outputs j with probability $\operatorname{Tr}(E_j\rho)$. In our setting of online learning of quantum states, a sequence of measurements $\mathcal{M}_t$ is given and each $\mathcal{M}_t$ can be described by a set of Hermitian matrices $\{E_{t,1}, E_{t,2}, \ldots, E_{t,K}\}$. In Appendix C.1, we show that our algorithm has the same regret upper bound on it for any K if we use the total variational norm as the loss function.

We also conduct numerical experiments for our algorithms about their dynamic and adaptive regret bounds on the k-shift and the path length settings. The results demonstrate that our algorithms improve upon non-adaptive algorithms when the quantum state is changing, confirming our theoretical results.

2 Preliminaries

In online convex optimization (refer to [27] for a comprehensive treatment), a player engages in a T-round game against an adversarial environment. In each round t, the player selects $x_t \in \mathcal{K}$, then the environment unveils the convex loss function ℓ_t, and the player incurs the loss $\ell_t(x_t)$. The player's objective is to minimize the (static) regret:

$$\text{Regret} = \sum_{t=1}^{T} \ell_t(x_t) - \min_{\varphi \in \mathcal{K}} \sum_{t=1}^{T} \ell_t(\varphi). \tag{3}$$

[2]Here we make no assumptions about how ℓ_t is revealed. In application, for example [1], they first obtain b_t by repeated measurements and then compute ℓ_t classically.

We can model the quantum tomography problem within the framework of online convex optimization: the convex domain $\mathcal{K}$ is taken to be C_n, the set of all trace-1 positive semidefinite (PSD) complex matrices of dimension 2^n:

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

The loss function can be $\ell_t(x) = |\mathrm{Tr}(E_t x_t) - \mathrm{Tr}(E_t \rho_t)|$ or $\ell_t(x) = (\mathrm{Tr}(E_t x_t) - \mathrm{Tr}(E_t \rho_t))^2$ for example, measuring the quality of approximating the ground-truth ρ_t.

Another goal is to bound the maximum number of mistakes we make during the learning process. We consider the case where $\forall t \in [T]$, b_t in the revealed loss function ℓ_t satisfies $|b_t - \mathrm{Tr}(E_t \rho_t)| \leq \frac{1}{3}\epsilon$. We consider it a mistake if the algorithm outputs an x_t such that $|\mathrm{Tr}(E_t x_t) - \mathrm{Tr}(E_t \rho_t)| > \epsilon$.

Notations. Let $\|\cdot\|_1$ denote the nuclear norm, i.e., the ℓ_1 norm of the singular values of a matrix. Let $x \bullet y = \mathrm{Tr}(x^\dagger y)$ denote the inner trace product between x and y. Denote $\lambda_i(X)$ to be the i^{th} largest eigenvalue of a Hermitian matrix X. $X \succeq 0$ means that the matrix X is positive semidefinite. For any positive integer n, denote $[n] := \{0, 1, \ldots, n-1\}$. Denote $\otimes$ to be the tensor product, i.e., for matrices $A = (A_{i_1 j_1})_{i_1 \in [m_1], j_1 \in [n_1]} \in \mathbb{C}^{m_1} \times \mathbb{C}^{n_1}$ and $B = (B_{i_2 j_2})_{i_2 \in [m_2], j_2 \in [n_2]} \in \mathbb{C}^{m_2} \times \mathbb{C}^{n_2}$, we have $A \otimes B = (A_{i_1 j_1} \cdot B_{i_2 j_2})_{i \in [m_1 m_2], j \in [n_1 n_2]} \in \mathbb{C}^{m_1 m_2} \times \mathbb{C}^{n_1 n_2}$ where $i = i_1 m_2 + i_2$, $j = j_1 n_2 + j_2$. Throughout the paper, $\tilde{O}$ omits poly-logarithmic terms.

3 Minimizing Dynamic Regret

In this section, we consider the case that ρ_t may change over time. In particular, we do not assume the number of times that ρ_t changes, but instead consider the total change over time measured by the path length. To this end, we study dynamic regret, where the path length $\mathcal{P}$ of the comparator in nuclear norm is restricted:

$$\mathcal{P} = \sum_{t=1}^{T-1} \|\varphi_t - \varphi_{t+1}\|_1. \tag{4}$$

Note that the set of comparators $\{\varphi_t\}$ do not have to be the ground truth states $\{\rho_t\}$; they can be any states satisfying the path length condition. We propose an algorithm achieving an $O(L\sqrt{T(n + \log(T))\mathcal{P}})$ dynamic regret bound, which instantiates copies of OMD algorithms with different learning rates as experts, and uses the multiplicative weight algorithm to select the best expert. The need for an expert algorithm arises because to obtain the dynamic regret bound, the learning rate of the OMD algorithm has to be a function of the path length, which is unknown ahead of time. The main algorithm is given in Algorithm 1, and the expert OMD algorithm is given in Algorithm 2. The following theorem gives the dynamic regret guarantee:

Theorem 1. *Assume the path length $\mathcal{P} = \sum_{t=1}^{T-1} \|\varphi_t - \varphi_{t+1}\|_1 \geq 1$ and the loss ℓ_t is convex, L-Lipschitz, and maps to $[0, 1]$, the dynamic regret of Algorithm 1 is bounded by $O(L\sqrt{T(n + \log(T))\mathcal{P}})$.*

The proof of Theorem 1 is deferred to Appendix A and we discuss some technical challenges here. Ref. [28] gives an $O(\sqrt{T\mathcal{P}})$ dynamic regret bound for online convex optimization:

$$\text{Dynamic Regret} = \sum_{t=1}^{T} \ell_t(x_t) - \min_{\varphi_1, \ldots, \varphi_t \in \mathcal{K}} \sum_{t=1}^{T} \ell_t(\varphi_t) \tag{5}$$

Algorithm 1: Dynamic Regret for Quantum Tomography

Input: a candidate set of η, $S = \{2^{-k-1} \mid 1 \le k \le \log T\}$, constant α.
Initialize $\log T$ experts $A_1, \ldots, A_{\log T}$, where A_k is an instance of Algorithm 2 with $\eta = 2^{-k-1}$ for all $k \in [\log T]$.
Set initial weights $w_1(k) = \frac{1}{\log T}$ for all $k \in [\log T]$.
for $t = 1, \ldots, T$ **do**
    Predict $x_t = \frac{\sum_k w_t(k) x_t(k)}{\sum_k w_t(k)}$, where $x_t(k)$ is the output of the k-th expert.
    Observe loss function $\ell_t(\cdot)$.
    Update the weights as
$$w_{t+1}(k) = w_t(k) e^{-\alpha \ell_t(x_t(k))}.$$
    Send gradients $\nabla \ell_t(x_t(k))$ to each expert A_k.
end for

Algorithm 2: OMD for Quantum Tomography

Input: domain $\mathcal{K} = (1 - \frac{1}{T})C_n + \frac{1}{T 2^n} I$, step size $\eta < \frac{1}{2L}$.
Define $R(x) = \sum_{k=1}^{2^n} \lambda_k(x) \log \lambda_k(x)$, $\nabla R(x) := I + \log(x)$, and let B_R denote the Bregman divergence defined by R.
Set $x_1 = 2^{-n} I$, and y_1 to satisfy $\nabla R(y_1) = \mathbf{0}$.
for $t = 1, \ldots, T$ **do**
    Predict x_t and receive loss $\ell_t(\mathrm{Tr}(E_t x_t))$.
    Define $\nabla_t = \ell_t'(\mathrm{Tr}(E_t x_t)) E_t$, where $\ell_t'(y)$ is a subderivative of ℓ_t with respect to y.
    Update y_{t+1} such that $\nabla R(y_{t+1}) = \nabla R(x_t) - \eta \nabla_t$.
    Update $x_{t+1} = \mathrm{argmin}_{x \in \mathcal{K}} B_R(x || y_{t+1})$.
end for

where $\sum_{t=1}^{T-1} \|\varphi_t - \varphi_{t+1}\|_2 = \mathcal{P}$. Directly adopting their result will lead to an **exponential dependence on** n, because the original algorithm uses ℓ_2 norm to measure the path length, which makes dual norm of the gradient depends polynomially on the dimension (2^n in this case). To avoid this issue, we use OMD with von Neumann entropy regularization to obtain a poly-logarithmic dependence on the dimension. From [5], we know that with the von Neumann entropy as the regularizer, the regret can be bounded by problem parameters measured in nuclear norm and spectral norm, which tend to be small for online shadow tomography. The analysis is in a similar vein as the classical OMD proof [27], but classical results often rely on theorems from calculus that cannot be straightforwardly applied to the complex domain.

Given the challenge of complex numbers, we use tools from functional analysis and quantum information theory to study the convergence of the algorithm. Our analysis deviates from the classical analysis in the following aspects:

- We use the notion of Gateaux differentiability and the Gateaux derivative to extend classical gradient-based methods to the quantum domain. In particular, we use the Gateaux derivative to define the Bregman divergence, a key ingredient of the OMD analysis. Subsequently, we show that the Bregman divergence obeys an Euler's inequality that corresponds to the generalized Pythagorean theorem in the classical analysis [27].

- To give a quantitative bound on the Bregman divergence between the current iterate

and the dual update, we cannot rely on the interpretation of the Bregman divergence as a notion of distance. Instead, we upper bound the quantity using the power series of the matrix exponential, and the Golden-Thompson inequality.

- When bounding the dynamic regret, we need to bound the difference of von Neumann entropy between two quantum states. This can be seen as a type of continuity bound for the von Neumann entropy, and was first given by [29] and later sharpened by [22]. The bound is known as the Fannes-Audenaert inequality with numerous applications in quantum information theory [30, 31, 32, 33].

Mistake bound of the path length setting. We can obtain a mistake bound in this setting when we assume the ground truth ρ_t has a bounded path length.

Corollary 1. *Let $\{\rho_t\}$ be a sequence of n-qubit mixed state whose path length $P_T = \sum_{t=1}^{T-1} \|\rho_{t+1} - \rho_t\|_1 \leq \mathcal{P}$, and let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements that are revealed to the learner one by one, each followed by a value $b_t \in [0,1]$ such that $|\operatorname{Tr}(E_t\rho_t) - b_t| \leq \frac{1}{3}\epsilon$. Then there is an explicit strategy for outputting hypothesis states $x_1, x_2, \ldots$ such that $|\operatorname{Tr}(E_t x_t) - \operatorname{Tr}(E_t\rho_t)| > \epsilon$ for at most $\tilde{O}(\frac{L^2 n\mathcal{P}}{\epsilon^2})$ times.*

Dynamic regret given dynamical models. So far, we have not made assumptions on how the state evolves. However, in many cases, the quantum state only evolves under a family of dynamical models. A natural question is the following: If we know the family of possible dynamical models in advance, can we achieve a better dynamic regret?

First, we consider the case when there is a fixed family of dynamical models. The most general linear maps $\Phi\colon C_n \to C_n$ implementable on quantum computers are *quantum channels* satisfying:

- Completely positive: For any $\rho \in C_n$ and m-dimensional identity matrix I_m, $\rho \otimes I_m \succeq 0$;
- Trace preserving: $\operatorname{Tr}(\Phi(\rho)) = \operatorname{Tr}(\rho) = 1$ for any quantum state $\rho \in C_n$.

A common example is $\Phi(x) = e^{iHt}xe^{-iHt}$, where H is a background Hamiltonian that evolves the quantum state following the Schrödinger equation.

We first consider the case with M possible dynamical models $\{\Phi^{(1)}, \ldots, \Phi^{(M)}\}$, each of which is a quantum channel. Following [34] and [28], we restrict a variant of the path length of the comparator:

$$\mathcal{P}' = \min_{\Phi^{(i)} \in \{\Phi^{(1)}, \ldots, \Phi^{(M)}\}} \sum_{t=1}^{T-1} \|\varphi_{t+1} - \Phi^{(i)}(\varphi_t)\|_1. \tag{6}$$

Corollary 2. *Under the assumption that we have M possible dynamical models $\{\Phi^{(1)}, \ldots, \Phi^{(M)}\}$ and the path length $\mathcal{P}' = \min_{\Phi^{(i)} \in \{\Phi^{(1)}, \ldots, \Phi^{(M)}\}} \sum_{t=1}^{T-1} \|\varphi_{t+1} - \Phi^{(i)}(\varphi_t)\|_1 \geq 1$, there is an algorithm with dynamic regret $O(L\sqrt{T(n + \log(T))\mathcal{P}'} + \sqrt{T \log(M \log(T))})$.*

In general, the possible dynamical models can be a continuous set. Considering that current quantum computers [23, 24, 25] commonly use 2D chips where a quantum gate only influences a constant number of qubits, it is natural to consider l-local quantum channels where each channel only influences at most l qubits (and perform identity on the other $n - l$ qubits), and the following result holds. The proofs of Corollary 2 and Corollary 3 are deferred to Appendix B.

Corollary 3. *Under the assumption that the dynamical models are l-local quantum channels, there is an algorithm with dynamic regret $\tilde{O}(L\sqrt{nT\mathcal{P}'} + 2^{6l}\sqrt{T})$, where $\mathcal{P}' = \min_{\Phi \in S_{n,l}} \sum_{t=1}^{T-1} \|\varphi_{t+1} - \Phi(\varphi_t)\|_1 \geq 1$ and $S_{n,l}$ is the set of all l-local quantum channels.*

4 Minimizing Adaptive Regret

In this section, we minimize the adaptive regret in Eq. (2) in the non-realizable case, i.e., b_t can be arbitrarily selected. We follow a meta-algorithm called Coin Betting for Changing Environment (CBCE) proposed by [20] with its black-box algorithm $\mathcal{B}$ being the Regularized Follow-the-Leader based algorithm introduced by [5] for learning quantum states. The algorithm is given in Algorithm 3.

Algorithm 3: Adaptive Regret for Quantum Tomography

1: **Input:** CBCE from [20] as the meta-algorithm, RFTL from [5] as the black-box algorithm.
2: Initialize experts $\mathcal{B}_J$, where $\mathcal{B}_J$ is an instance of the RFTL algorithm over a different interval J, $J \in \mathcal{J}$ where $\mathcal{J}$ is the set of geometric intervals.
3: Initialize a prior distribution over $\mathcal{B}_J$.
4: **for** $t = 1, \ldots, T$ **do**
5: Predict among the black-box experts $\mathcal{B}_J$ according to CBCE.
6: Update the distribution over the black-box experts $\mathcal{B}_J$ according to CBCE.
7: **end for**

The basic idea is that the CBCE meta-algorithm can transfer the regret bound of any black-box algorithm to strongly adaptive regret, and the RFTL algorithm with von Neumann entropy regularization achieves $O(\sqrt{T})$ regret bound which can serve as the black-box algorithm. Formally:

Lemma 1. *Let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements on an n-qubit state presented to the learner, and suppose the loss ℓ_t is convex, L-Lipschitz, and maps to $[0,1]$. Then Algorithm 3 guarantees strongly adaptive regret SA-Regret$_T^{\mathcal{A}}(\tau) = O(L\sqrt{n\tau \log(T)})$ for all T.*

k-shift regret bound. As an application of Lemma 1, we derive a k-shift regret bound. In the k-shift setting, we assume that φ can change at most k times, and we derive the following theorem.

Theorem 2. *Under the same setting as Lemma 1, if we allow φ to change at most k times, we have an $O(L\sqrt{knT \log(T)})$ bound on the k-shift regret $R_{k\text{-shift}}^{\mathcal{A}}$:*

$$R_{k\text{-shift}}^{\mathcal{A}} = \sum_{t=1}^{T} \ell_t(E_t x_t^{\mathcal{A}}) - \min_{1=t_1 \le t_2 \le \cdots \le t_{k+1}=T+1} \min_{\varphi_1, \varphi_2, \ldots, \varphi_k \in C_n} \sum_{j=1}^{k} \sum_{t=t_j}^{t_{j+1}-1} \ell_t(E_t \varphi_j). \tag{7}$$

Mistake bound of the k-shift setting. With Theorem 2 in hand, when we further assume that the ground truth ρ_t also shifts at most k times, we can derive the following mistake bound.

Corollary 4. *Let $\{\rho_t\}$ be a sequence of n-qubit mixed state which changes at most k times, and let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements that are revealed to the learner one by one, each followed by a value $b_t \in [0,1]$ such that $|\operatorname{Tr}(E_t\rho_t) - b_t| \le \frac{1}{3}\epsilon$. Then there is an explicit strategy for outputting hypothesis states $x_1, x_2, \ldots$ such that $|\operatorname{Tr}(E_t x_t) - \operatorname{Tr}(E_t \rho_t)| > \epsilon$ for at most $O\left(\frac{kn}{\epsilon^2} \log\left(\frac{kn}{\epsilon^2}\right)\right)$ times.*

The proofs of the claims in this section are deferred to Appendix C.

5 Numerical Experiments

In this section, we present numerical results which support our regret bounds.[3] Specifically, we conduct experiments for comparing with non-adaptive algorithms and testing our performance of the k-shift setting and the path length setting, respectively. All the experiments are performed on Visual Studio Code 1.6.32 on a 2021 MacBook Pro (M1 Pro chip) with 10-core CPU and 16GB memory. We generate random n-qubit quantum states using the QuTiP package [35, 36]. In the k-shift setting, each of the k changes of the quantum state are chosen uniformly at random from all iterations $\{1, 2, \ldots, T\}$. In the path length setting, our quantum state evolves slowly following a background Hamiltonian H also generated by QuTiP. At time t we generate a POVM measurement E_t with eigenvalues and eigenvectors chosen uniformly at random from $[0, 1]$ and ℓ_2-norm unit vectors, respectively. Then, we apply the measurement, receive loss, and update our prediction using Algorithm 1 and Algorithm 3. For all experiments, the instantaneous regret at each time step is computed using the ground truth state, which changes over each run depending on the setting.

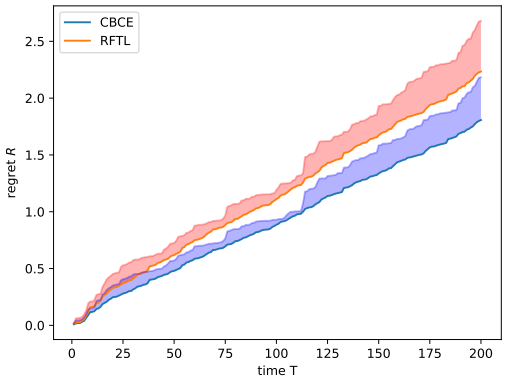
Comparison to non-adaptive algorithms. One eminent strength of adaptive online learning of quantum states is its ability to capture the changing states. Previous algorithms such as regularized follow-the-leader (RFTL) perform well for predicting fixed quantum states [5], but they are not adaptive when the quantum state evolves. We compare the cumulative regret produced by RFTL and our algorithms in the k-shift setting and the path length setting. The results are illustrated as follows.

The advantage of our algorithms in the k-shift setting is presented in Figure 1. In Subfigure(a), we plot the long term changes of the cumulative regret. The two curves have similar trends but CBCE achieves smaller regret. In Subfigure(b)-(f), we choose a small time scale $T = 20$ to display the details. It is clearly shown that the distances between two curves grow larger as T increases, which means our algorithm maintains the advantage in the long run.

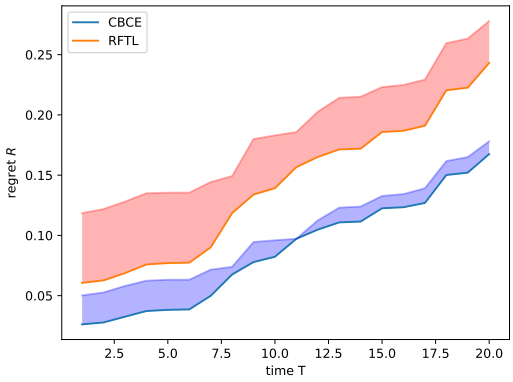
Figure 2 shows the advantage of our algorithms in the path length setting. In Subfigure(a) and (b), we plot the comparison between our Algorithm 1 (denoted by "DOMD") and the non-adaptive algorithm RFTL with different learning rates. It can be seen that in a single run, our algorithm attains the minimum regret and is comparable to RFTL with the best learning rate. Indeed, given the similarity of RFTL and OMD, we expect RFTL with the optimal learning rate to do as well as DOMD in practice. However, since the optimal learning rate for OMD depends on the path length, it is unknown a priori, and non-adaptive methods require more tuning to achieve good performance. In Subfigure(c) and (d), we plot the comparison between our Algorithm 3 (denoted by "CBCE") and the non-adaptive RFTL with different η. There is a significant gap between the regret of CBCE and RFTL with the best learning rate from the grid search, demonstrating the interesting phenomenon that empirically, CBCE with adaptive regret guarantees also does well in the path length setting.

Adaptive regret bound on the k-shift setting. We also conduct experiments to verify our adaptive regret bound $O(\sqrt{knT \log T})$ in the k-shift setting (Theorem 2). Results are shown in Figure 3.

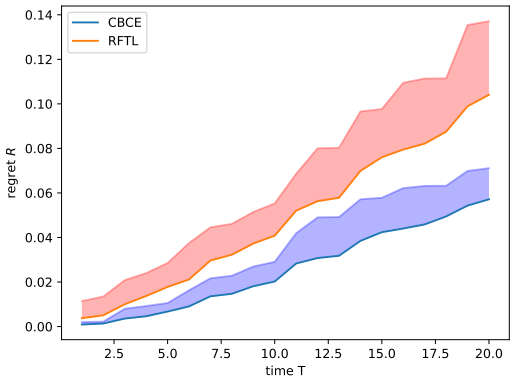
[3]The implementation codes can be found at the link: https://github.com/EstelYang/Adaptive-Online-Learning-of-Quantum-States



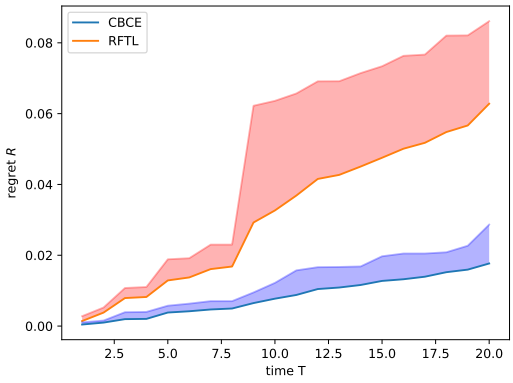
(a) k-shift setting, $k = 80$, $n = 2$.



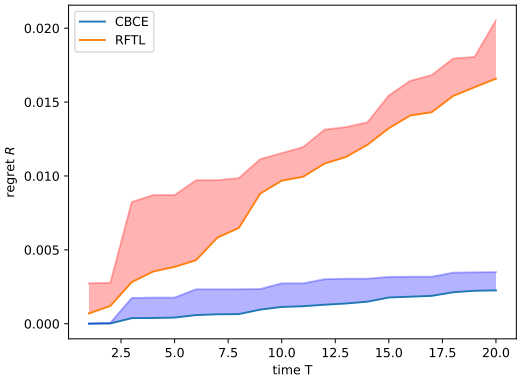
(b) k-shift setting, $k = 4$, $n = 2$.



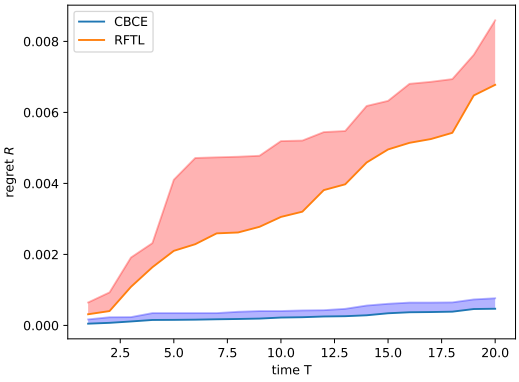
(c) k-shift setting, $k = 4$, $n = 3$.



(d) k-shift setting, $k = 4$, $n = 4$.

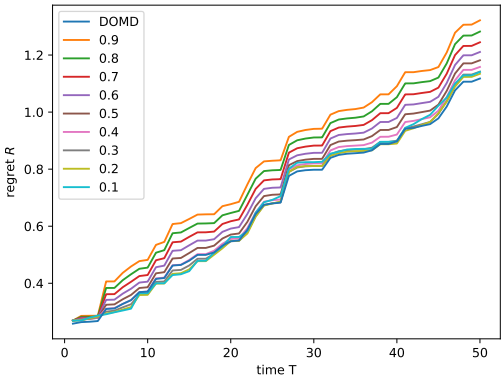


(e) k-shift setting, $k = 4$, $n = 5$.

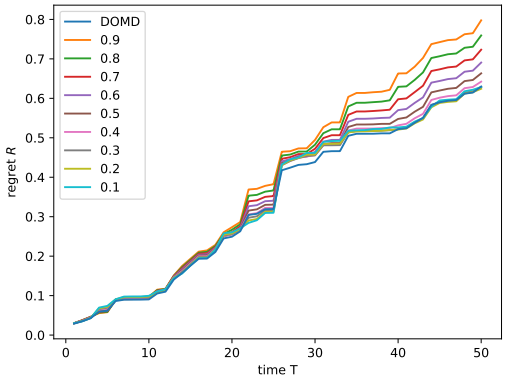


(f) k-shift setting, $k = 4$, $n = 6$.

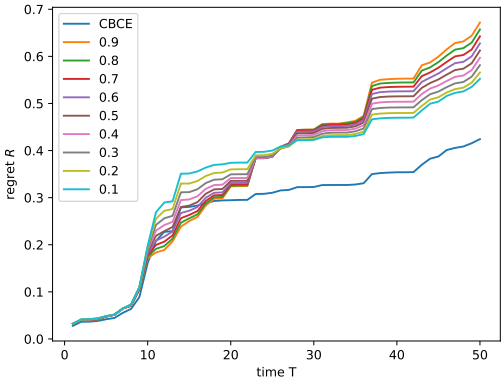
Figure 1: Comparison between Algorithm 3 and non-adaptive RFTL in the k-shift setting. The regret attained by our adaptive Algorithm 3 is denoted by "CBCE" (the blue curve with purple shadow), and the regret generated by the non-adaptive algorithm RFTL is denoted by the orange curve with red shadow. The lower solid line stands for the average value of regret over 10 random experiments and the upper line stands for the maximum.



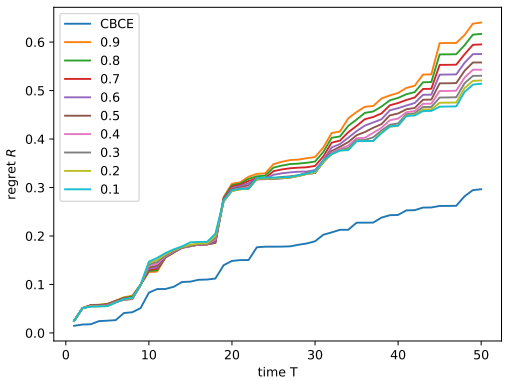
(a) Path length setting, Algorithm 1, $n = 2$.



(b) Path length setting, Algorithm 1, $n = 3$.

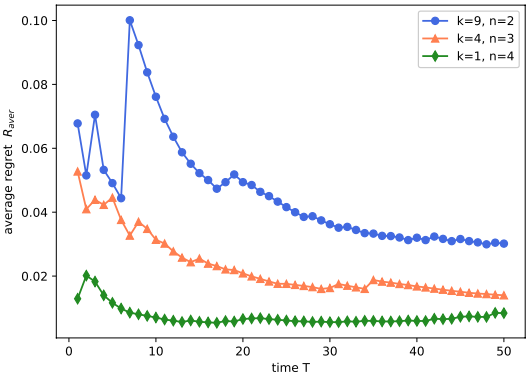


(c) Path length setting, Algorithm 3, $n = 2$.

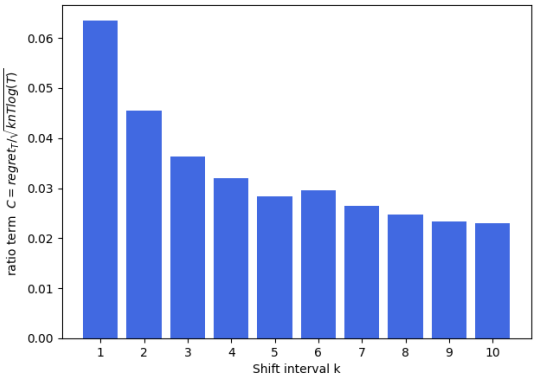


(d) Path length setting, Algorithm 3, $n = 3$.

Figure 2: Comparison between our algorithms and non-adaptive RFTL in the path length setting. Our algorithms are denoted by "DOMD" / "CBCE", and the non-adaptive algorithm RFTL with different learning rates η are denoted by the value of η.



(a) Average Regret R_{aver}



(b) Ratio $C = R_{k\text{-shift}}/\sqrt{knT\log T}$

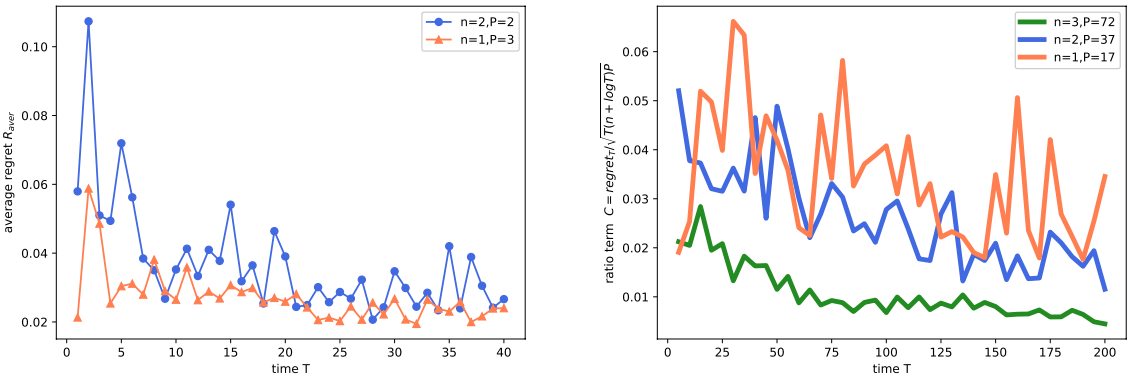
Figure 3: Adaptive regret bounds in the k-shift setting.

In Subfigure(a), we plot the average regret $R_{\text{aver}} = \text{regret}_T/T$ as T grows. Every datapoint was executed by 100 random experiments and taking the maximum regret, since

regret is an upper bound. Under different parameter choices (k, n), R_{aver} always converges to 0 when T increases, implying a sublinear regret $R_{k\text{-shift}}$. More concretely, curves with larger k has larger R_{aver} and converges more slowly, consistent with the intuition that more shifts make the learning task more difficult.

In Subfigure(b), to verify the precise bound we plot the ratio $C = \text{regret}_T/\sqrt{knT\log T}$. We choose $n = 2, k = 10, T = 200$, and repeat the experiments 50 times randomly. We then plot the maximum of the regret constant C on every interval. As the figure shows, the ratio is limited by a constant $C \leq 0.07$. This strongly supports our upper bound $O(\sqrt{knT\log T})$ on regret_T with every fixed n.

Dynamic regret bound on the path length setting. In addition, we conduct experiments in terms of average regret R_{aver} and ratio term $C_{\text{path}} = R_{\text{path}}/\sqrt{T(n+\log T)\mathcal{P}}$ to testify our dynamic regret bound $O(L\sqrt{T(n+\log T)\mathcal{P}})$ in the path length setting (Theorem 1). Results are shown in Figure 4.



(a) Average Regret R_{aver}

(b) Ratio $C = R_{\text{path}}/\sqrt{T(n+\log T)\mathcal{P}}$

Figure 4: Dynamic regret bound on the path length setting.

In Subfigure(a), we plot the change of average regret $R_{\text{aver}} = \text{regret}_T/T$. We observe that as T grows, R_{aver} converges to 0 with different pairs of $(n, \mathcal{P})$, indicating that R_{path} is sublinear with respect to T. In addition, the blue curve representing the average regret of a 2-qubit quantum state appears higher than the orange curve representing the average regret of a 1-qubit quantum state, consistent with the fact that our regret bound depends polynomially on the number of qubits.

In Subfigure(b), we choose different scales of T to demonstrate how the ratio C evolves in the long term. For every fixed T, we execute 100 random experiments and take the maximum ratio term C. Although there are noticeable fluctuations, it can be seen that the curve is non-increasing and only fluctuates within a fixed interval of 0 to 0.065, which shows that C has a constant upper bound $C' = 0.065$. In conclusion, $R_{\text{path}}/\sqrt{T(n+\log T)\mathcal{P}}$ is within the range of a small positive constant.

6 Conclusions

In this paper, we consider online learning of changing quantum states. We give algorithms with adaptive and dynamic regret bounds that are polynomial in the number of qubits and sublinear in the number of measurements, and verify our results with numerical experiments.

Our work leaves a couple of open questions for future investigation:

- Lower bound in T and path length P are known in [28], but there is no lower bound in n to our knowledge. Can we prove an $\Omega(\sqrt{n})$ lower bound for dynamic regret or adaptive regret of online learning of quantum states? We noticed a recent work [37] that proposed a lower bound in terms of n for quantum channels. It would be intriguing in future work to explore whether these techniques can be adapted to the quantum state scenario.

- Although our algorithms do not care how the loss function is revealed, in practice, a standard way is to measure several copies of the state at each step, so the copies need to follow the same dynamics. Accounting for different state evolution for different state copies is left as future work.

Acknowledgments

TL, XW, and RY were supported by the National Natural Science Foundation of China (Grant Numbers 92365117 and 62372006), and The Fundamental Research Funds for the Central Universities, Peking University.

References

[1] Scott Aaronson. "Shadow tomography of quantum states". SIAM Journal on Computing **49**, STOC18–368–STOC18–394 (2020).

[2] Costin Bădescu and Ryan O'Donnell. "Improved quantum data analysis". In Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing. Pages 1398–1411. (2021).

[3] Fernando G.S.L. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu. "Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning". In Proceedings of the 46th International Colloquium on Automata, Languages, and Programming. Volume 132 of Leibniz International Proceedings in Informatics, pages 27:1–27:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2019).

[4] Scott Aaronson and Guy N Rothblum. "Gentle measurement of quantum states and differential privacy". In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing. Pages 322–333. (2019).

[5] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. "Online learning of quantum states". Advances in neural information processing systems**31** (2018). url: https://proceedings.neurips.cc/paper_files/paper/2018/file/c1a3d34711ab5d85335331ca0e57f067-Paper.pdf.

[6] Feidiao Yang, Jiaqing Jiang, Jialin Zhang, and Xiaoming Sun. "Revisiting online quantum state learning". In Proceedings of the AAAI Conference on Artificial Intelligence. Volume 34, pages 6607–6614. (2020).

[7] Yifang Chen and Xin Wang. "More practical and adaptive algorithms for online quantum state learning" (2020). arXiv:2006.01013

[8] Josep Lumbreras, Erkka Haapasalo, and Marco Tomamichel. "Multi-armed quantum bandits: Exploration versus exploitation when learning properties of quantum states". Quantum **6**, 749 (2022).

[9] Julian Zimmert, Naman Agarwal, and Satyen Kale. "Pushing the efficiency-regret pareto frontier for online learning of portfolios and quantum states". In Conference on Learning Theory. Pages 182–226. PMLR (2022). url: https://proceedings.mlr.press/v178/zimmert22a.html.

[10] Harrison Ball, Thomas M. Stace, Steven T. Flammia, and Michael J. Biercuk. "Effect of noise correlations on randomized benchmarking". Physical Review A **93**, 022303 (2016).

[11] Daniel Greenbaum and Zachary Dutton. "Modeling coherent errors in quantum error correction". Quantum Science and Technology **3**, 015007 (2017).

[12] Richard Kueng, David M. Long, Andrew C. Doherty, and Steven T. Flammia. "Comparing experiments to the fault-tolerance threshold". Physical Review Letters **117**, 170502 (2016).

[13] Joel J. Wallman. "Bounding experimental quantum error rates relative to fault-tolerant thresholds" (2015). arXiv:1511.00727

[14] Joel Wallman, Chris Granade, Robin Harper, and Steven T. Flammia. "Estimating the coherence of noise". New Journal of Physics **17**, 113020 (2015).

[15] Quoc Hoan Tran and Kohei Nakajima. "Learning temporal quantum tomography". Physical Review Letters **127**, 260401 (2021).

[16] John Preskill. "Quantum computing in the NISQ era and beyond". Quantum **2**, 79 (2018).

[17] Martin Zinkevich. "Online convex programming and generalized infinitesimal gradient ascent". In International Conference on Machine Learning. Pages 928–936. (2003). url: https://dl.acm.org/doi/10.5555/3041838.3041955.

[18] Elad Hazan and Comandur Seshadhri. "Efficient learning algorithms for changing environments". In International Conference on Machine Learning. Pages 393–400. (2009).

[19] Amit Daniely, Alon Gonen, and Shai Shalev-Shwartz. "Strongly adaptive online learning". In International Conference on Machine Learning. Pages 1405–1411. PMLR (2015). url: https://proceedings.mlr.press/v37/daniely15.html.

[20] Kwang-Sung Jun, Francesco Orabona, Stephen Wright, and Rebecca Willett. "Improved strongly adaptive online learning using coin betting". In Artificial Intelligence and Statistics. Pages 943–951. PMLR (2017). url: https://proceedings.mlr.press/v54/jun17a.html.

[21] Ashok Cutkosky. "Parameter-free, dynamic, and strongly-adaptive online learning". In International Conference on Machine Learning. Pages 2250–2259. PMLR (2020). url: https://proceedings.mlr.press/v119/cutkosky20a.html.

[22] Koenraad M. R. Audenaert. "A sharp continuity estimate for the von Neumann entropy". Journal of Physics A: Mathematical and Theoretical **40**, 8127–8136 (2007).

[23] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, and et al. "Quantum supremacy using a programmable superconducting processor". Nature **574**, 505–510 (2019).

[24] IBM Quantum. "Pushing quantum performance forward with our highest quantum volume yet". https://research.ibm.com/blog/quantum-volume-256 (2022). Accessed: 2022-05-13.

[25] Qingling Zhu, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, and et al. "Quantum computational advantage via 60-qubit 24-cycle random circuit sampling". Science Bulletin **67**, 240–245 (2022).

[26] Weiyuan Gong and Scott Aaronson. "Learning distributions over quantum measurement outcomes". In International Conference on Machine Learning. Pages 11598–11613. PMLR (2023). url: https://proceedings.mlr.press/v202/gong23a.html.
[27] Elad Hazan. "Introduction to online convex optimization". Foundations and Trends® in Optimization **2**, 157–325 (2016).
[28] Lijun Zhang, Shiyin Lu, and Zhi-Hua Zhou. "Adaptive online learning in dynamic environments". In Advances in Neural Information Processing Systems. Volume 31. (2018). arXiv:1810.10815.
[29] M. Fannes. "A continuity property of the entropy density for spin lattice systems". Communications in Mathematical Physics **31**, 291–294 (1973).
[30] Michael M. Wolf, Frank Verstraete, Matthew B. Hastings, and J. Ignacio Cirac. "Area laws in quantum systems: mutual information and correlations". Physical Review Letters **100**, 070502 (2008).
[31] Johan Åberg. "Catalytic coherence". Physical Review Letters **113**, 150402 (2014).
[32] Andreas Winter and Dong Yang. "Operational resource theory of coherence". Physical Review Letters **116**, 120404 (2016).
[33] Eric Chitambar and Gilad Gour. "Quantum resource theories". Reviews of Modern Physics **91**, 025001 (2019).
[34] Eric Hall and Rebecca Willett. "Dynamical models and tracking regret in online convex programming". In International Conference on Machine Learning. Pages 579–587. PMLR (2013). url: https://proceedings.mlr.press/v28/hall13.html.
[35] J.R. Johansson, P.D. Nation, and Franco Nori. "QuTiP: An open-source Python framework for the dynamics of open quantum systems". Computer Physics Communications **183**, 1760–1772 (2012).
[36] J.R. Johansson, P.D. Nation, and Franco Nori. "QuTiP 2: A Python framework for the dynamics of open quantum systems". Computer Physics Communications **184**, 1234–1240 (2013).
[37] Asad Raza, Matthias C. Caro, Jens Eisert, and Sumeet Khatri. "Online learning of quantum processes" (2024). arXiv:2406.04250.
[38] Heinz H. Bauschke and Patrick L. Combettes. "Convex analysis and monotone operator theory in hilbert spaces". Springer Publishing Company, Incorporated. (2011). 1st edition.
[39] Eric P. Hanson and Nilanjana Datta. "Maximum and minimum entropy states yielding local continuity bounds". Journal of Mathematical Physics **59**, 042204 (2018).
[40] Joel A. Tropp. "From joint convexity of quantum relative entropy to a concavity theorem of lieb". Proceedings of the American Mathematical Society **140**, 1757–1760 (2012). url: http://www.jstor.org/stable/41505631.
[41] Rajendra Bhatia. "Matrix analysis". Volume 169. Springer. New York (1997).
[42] Göran Lindblad. "Completely positive maps and entropy inequalities". Communications in Mathematical Physics **40**, 147–151 (1975).
[43] William F. Stinespring. "Positive functions on c*-algebras". Proceedings of the American Mathematical Society **6**, 211–216 (1955). url: https://community.ams.org/journals/proc/1955-006-02/S0002-9939-1955-0069403-4/S0002-9939-1955-0069403-4.pdf.
[44] Emanuel Knill. "Approximation by quantum circuits" (1995). arXiv:quant-ph/9508006

A Proof of Dynamic Regret Bounds for Quantum Tomography

We prove Theorem 1 in this section.

We use the online mirror descent (OMD) algorithm (similar to Algorithm 1 in [5]) to replace the online gradient descent (OGD) expert in [28]. We denote G_R to be an upper bound on the spectral norm of $\nabla_t = \ell_t'(\mathrm{Tr}(E_t x_t))E_t$, and $D_R = \sqrt{\max_{x,y\in\mathcal{K}}\{R(x) - R(y)\}}$ to be the diameter of $\mathcal{K}$ with respect to the function R. The main motivation is that the dimension now is 2^n and we need logarithmic dependence on dimension for our regret bound, where ℓ_2 norm (used by OGD) is not the most appropriate choice and fails to control G_R.

Instead, the OMD algorithm with von Neumann entropy has a polynomial upper bound on both D_R and G_R, and the only question left is how to incorporate the path length into the analysis of OMD to achieve dynamic regret. We observe that such incorporation is natural, due to the essential similarity between OGD and OMD. We bound the path distance $\|\varphi_t - \varphi_{t+1}\|$ in nuclear norm, and bound the gradient of the regularization $\nabla R(x)$ by spectral norm.

A potential issue is that the von Neumann entropy contains logarithmic terms, which might cause the gradient to explode when λ_k is very small. To mitigate this issue, we mix x with $\frac{I}{T2^n}$: $\tilde{x} = (1 - \frac{1}{T})x + \frac{I}{T2^n}$, for both x_t and φ_t. Notice that this mixing does not hurt the regret bound: mixing such a small fraction of the identity only increases the regret by $O(LT\frac{1}{T}) = O(L)$, since the losses are L-Lipschitz. In addition, the path length will only decrease, since we are working over a smaller domain. Henceforth we will consider the modified $\mathcal{K} = (1 - \frac{1}{T})C_n + \frac{1}{T2^n}I$ throughout the proof.

Our proof uses tools in convex analysis for Hilbert spaces. Since all matrices used in Algorithm 2 are Hermitian (by Proposition 1), it suffices to consider the Hilbert space of Hermitian matrices equipped with the trace inner product. Note that this is a real Hilbert space: the vector space of Hermitian matrices only allows scalar multiplication over the field of reals.

We first present some basic results for the von Neumann Entropy. For completeness, we give the following definitions from [38]; for a more comprehensive treatment of the topic of analysis in Hilbert spaces, see also [38].

Definition 1 (Gateaux differentiability)**.** *Let C be a subset of a Hilbert space $\mathcal{H}$, and $f : C \to [-\infty, \infty]$. Let $x \in C$ such that for each $y \in \mathcal{H}$, $\exists \alpha > 0$ such that $[x, x + \alpha y] \subset C$. Then f is Gateaux differentiable at x if there exists a bounded linear operator $Df(x)$ called the Gateaux derivative of f at x, such that*

$$Df(x)y = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha y) - f(x)}{\alpha}.$$

Definition 2 (Directional derivatives)**.** *For a Hilbert space $\mathcal{H}$, let $f : \mathcal{H} \to (-\infty, \infty]$ be proper, and $x \in domf, y \in \mathcal{H}$. The directional derivative of f at x in the direction y is*

$$f'(x; y) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha y) - f(x)}{\alpha},$$

if the limit exists on the extended real line.

It is well-known that f is Gateaux differentiable at x if and only if all directional derivatives at x exist and they form a bounded linear operator. By the Riesz-Frechet representation theorem, let $\langle \cdot, \cdot \rangle$ denote the Hilbert space inner product, there exists a unique $\nabla f(x) \in \mathcal{H}$ such that for each $y \in \mathcal{H}$,

$$f'(x; y) = \langle y, \nabla f(x) \rangle.$$

Lemma 2 (Lemma VI.4 in Ref. [39])**.** *The von Neumann entropy $R(x) = \mathrm{Tr}(x \log x)$ for Hermitian PD matrices x is Gateaux differentiable, and the Gateaux derivative is*

$$\nabla R(x) = \log x + I.$$

The quantum analog of the classical Bregman divergence is therefore defined as

$$B_R(x\|y) = R(x) - R(y) - \nabla R(y) \bullet (x - y).$$

The above expression is also known as the quantum relative entropy of x with respect to y, the quantum information divergence, and the von Neumann divergence [40].

Proposition 1. *The variable y_t in Algorithm 2 is always Hermitian PD.*

Proof. By the definition of y_t and $\nabla R(\cdot)$, we have $\log(y_{t+1}) = \log(x_{t-1}) - \eta \nabla_{t-1}$. Recall that the loss functions are of the form $\ell_t(\rho) = \ell(E_t \bullet \rho)$. Since E_t is Hermitian, $\nabla_{t-1} = \ell'(\cdot)E_{t-1}$ is also Hermitian. The lemma follows by definition. □

Proposition 2. *For all Hermitian PSD matrices x, y, $B_R(x\|y) \in \mathbb{R}$. For all $x \in \mathcal{K}$ and all Hermitian PD y, $B_R(x\|y) \geq 0$, and for all $\varphi_1 \in \mathcal{K}$ and x_1 as defined in Algorithm 2, $B_R(\varphi_1\|x_1) \leq \max_{x,y\in\mathcal{K}} \{R(x) - R(y)\}$.*

Proof. The first claim follows from the definition of R and ∇R. Since both x, y are both positive definite, by Fact 3 in [40] (Klein's inequality), $B_R(x\|y) \geq 0$. For the last claim, $\varphi_1 \in \mathcal{K}$ be any element, to show $B_R(\varphi_1\|x_1) = R(\varphi_1) - R(x_1) - \nabla R(x_1) \bullet (\varphi_1 - x_1) \leq \max_{x,y\in\mathcal{K}} R(x) - R(y)$, we only need to show $\nabla R(x_1) \bullet (\varphi_1 - x_1) \geq 0$. By direct computation,

$$\nabla R(x_1) \bullet (\varphi_1 - x_1) = (I + \log(2^{-n} I)) \bullet (\varphi_1 - 2^{-n} I) = (-n \log(2) I) \bullet (\varphi_1 - 2^{-n} I) = 0,$$

since $\mathrm{Tr}(\varphi_1) = \mathrm{Tr}(2^{-n} I) = 1$. □

Proposition 3 (Euler's inequality)**.** *Let x_t, y_t be as defined in Algorithm 2, then for any $\varphi \in \mathcal{K}$, we have*

$$B_R(\varphi\|x_t) \leq B_R(\varphi\|y_t).$$

Proof. For $x \in \mathcal{K}$,

$$B_R(x\|y) = \mathrm{Tr}(x \log x - x \log y - x + y).$$

By Lemma 2, $\mathrm{Tr}(x \log x)$ is Gateaux differentiable on the set of positive definite Hermitian matrices. Since $\mathrm{Tr}(x \log y)$ and $\mathrm{Tr}(x)$ are both Gateaux differentiable, $B_R(x\|y)$ is also Gateaux differentiable as a function of x. By definition, $x_t = \mathrm{argmin}_{x\in\mathcal{K}} B_R(x\|y_t)$, and therefore for $\alpha \in (0, 1]$ and any $\varphi \in \mathcal{K}$,

$$B_R(x_t\|y_t) \leq B_R((1 - \alpha)x_t + \alpha\varphi\|y_t).$$

Dividing by α and taking $\alpha \downarrow 0$, we have

$$\nabla B_R(x_t\|y_t) \bullet (\varphi - x_t) \geq 0.$$

Expanding the derivative, $(\nabla R(x_t) - \nabla R(y_t)) \bullet (\varphi - x_t) \geq 0$. Observe that we have the decomposition

$$(\nabla R(y_t) - \nabla R(x_t)) \bullet (\varphi - x_t) = B_R(x_t||y_t) - B_R(\varphi||y_t) + B_R(\varphi||x_t),$$

which means $B_R(\varphi||y_t) \geq B_R(x_t||y_t) + B_R(\varphi||x_t) \geq B_R(\varphi||x_t)$, because $B_R(x_t||y_t) \geq 0$ by Proposition 2. □

We now prove the main lemma which gives a dynamic regret bound of the OMD algorithm, when the learning rate η is chosen optimally.

Lemma 3. *Assume the path length $\mathcal{P} = \sum_{t=1}^{T-1} \|\varphi_t - \varphi_{t+1}\|_1 \geq 1$ where $\|\cdot\|_1$ is the nuclear norm, the dynamic regret of Algorithm 2 is upper bounded by $O(L\sqrt{T(n+\log T)\mathcal{P}})$, if we choose $\eta = \sqrt{\frac{\mathcal{P}(n+\log T)}{TL^2}}$.*

Proof. We follow the standard analysis of OMD, assuming ℓ_t are L-Lipschitz. By convexity of loss ℓ_t, we have that

$$\ell_t(\mathrm{Tr}(E_t x_t)) - \ell_t(\mathrm{Tr}(E_t \varphi_t)) \leq \ell_t'(\mathrm{Tr}(E_t x_t))(\mathrm{Tr}(E_t x_t) - \mathrm{Tr}(E_t \varphi_t)) = \nabla_t \bullet (x_t - \varphi_t). \tag{8}$$

The following property of Bregman divergences follows from the definition: for any Hermitian PSD matrices x, y, z,

$$(x - y) \bullet (\nabla R(z) - \nabla R(y)) = B_R(x\|y) - B_R(x\|z) + B_R(y\|z). \tag{9}$$

Now, take $x = \varphi_t$, $y = x_t$, and $z = y_{t+1}$ and combine (8) and (9). Writing $\ell_t(x) = \ell_t(\mathrm{Tr}(E_t x))$, we have

$$\begin{aligned}
&\ell_t(x_t) - \ell_t(\varphi_t) \\
\leq &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_t\|y_{t+1}) + B_R(x_t\|y_{t+1})) \\
\leq &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_t\|x_{t+1}) + B_R(x_t\|y_{t+1})) \\
= &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_{t+1}\|x_{t+1}) + B_R(\varphi_{t+1}\|x_{t+1}) - B_R(\varphi_t\|x_{t+1}) + B_R(x_t\|y_{t+1})),
\end{aligned}$$

where the second inequality is due to Proposition 3.

The following lemma gives an upper bound on one of the quantities in Eq. (10), which we will use next in the proof. In vanilla OMD, this quantity is upper bounded using the fact that the Bregman divergence can be interpreted as a distance measured by a certain local norm; however, under complex domain, this fact is no longer immediate from the Mean Value Theorem. We instead prove the upper bound by using properties of the dual space updates y_{t+1}.

Lemma 4. *Let G_R be an upper bound on $\|\nabla_t\|$, then $B_R(x_t\|y_{t+1}) \leq \frac{1}{2}\eta^2 G_R^2$.*

Proof. First, note that by definition, y_{t+1} satisfies

$$\log(y_{t+1}) = \log(x_t) - \eta\nabla_t. \tag{10}$$

Then we have $y_{t+1} = \exp(\log(x_t) - \eta\nabla_t)$, and by the Golden-Thompson inequality,

$$\mathrm{Tr}(y_{t+1}) \leq \mathrm{Tr}(\exp(\log(x_t))\exp(-\eta\nabla_t)) \leq \mathrm{Tr}(x_t \exp(-\eta\nabla_t)). \tag{11}$$

Expanding the Bregman divergence, we have

$$\begin{aligned}
B_R(x_t\|y_{t+1}) &= \mathrm{Tr}(x_t\log(x_t)) - \mathrm{Tr}(y_{t+1}\log(y_{t+1})) - (I + \log(y_{t+1})) \bullet (x_t - y_{t+1}) \\
&= \mathrm{Tr}(x_t\log(x_t)) - \mathrm{Tr}(x_t\log(y_{t+1})) - 1 + \mathrm{Tr}(y_{t+1}) \\
&= \mathrm{Tr}(x_t\log(x_t)) - \mathrm{Tr}(x_t(\log(x_t) - \eta\nabla_t)) - 1 + \mathrm{Tr}(y_{t+1}) && \text{(by (10))} \\
&= \eta\,\mathrm{Tr}(x_t\nabla_t) - 1 + \mathrm{Tr}(y_{t+1}) \\
&\leq \eta\,\mathrm{Tr}(x_t\nabla_t) - 1 + \mathrm{Tr}(x_t\exp(-\eta\nabla_t)). && \text{(by (11))}
\end{aligned}$$

By definition, the matrix exponential can be written as an infinite sum,

$$\exp(-\eta\nabla_t) = \sum_{k=0}^{\infty} \frac{1}{k!}(-\eta\nabla_t)^k.$$

Using this expression in the last term of the inequality,

$$\operatorname{Tr}(x_t \exp(-\eta\nabla_t)) = 1 - \eta \operatorname{Tr}(x_t\nabla_t) + \frac{1}{2}\eta^2 \operatorname{Tr}(x_t\nabla_t^2) + \sum_{k=3}^{\infty} \frac{1}{k!} \operatorname{Tr}(x_t(-\eta\nabla_t)^k).$$

Now we proceed to show that $\sum_{k=3}^{\infty} \frac{1}{k!} \operatorname{Tr}(x_t(-\eta\nabla_t)^k) \le 0$ if η is sufficiently small.

$$\begin{aligned}\sum_{k=3}^{\infty} \frac{1}{k!} \operatorname{Tr}(x_t(-\eta\nabla_t)^k) &= \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \operatorname{Tr}(x_t(-\eta\nabla_t)^{2k+1}) + \frac{1}{(2k+2)!} \operatorname{Tr}(x_t(-\eta\nabla_t)^{2k+2}) \\ &= \sum_{k=1}^{\infty} \frac{\eta^{2k+1}}{(2k+1)!} \operatorname{Tr}\left(x_t\left(-\nabla_t^{2k+1} + \frac{\eta}{2k+2}\nabla_t^{2k+2}\right)\right).\end{aligned}$$

Let $\nabla_t = VQV^\dagger$ be the eigendecomposition of ∇_t. Then

$$-\nabla_t^{2k+1} + \frac{\eta}{2k+2}\nabla_t^{2k+2} = VQ^{2k+1}(-I + \frac{\eta}{2k+2}Q)V^\dagger.$$

Since $\eta < \frac{1}{2L}$, and $\|\nabla_t\| \le L$,

$$-I + \frac{\eta}{2k+2}Q \preceq 0, \text{and} \quad \operatorname{Tr}\left(x_t\left(-\nabla_t^{2k+1} + \frac{\eta}{2k+2}\nabla_t^{2k+2}\right)\right) \le 0.$$

Because the summands are nonpositive, we conclude that $\sum_{k=3}^{\infty} \frac{1}{k!} \operatorname{Tr}(x_t(-\eta\nabla_t)^k) \le 0$. The lemma follows from applying the generalized Cauchy-Schwartz inequality (see 5, Claim 14 and 41, Exercise IV.1.14, page 90),

$$B_R(x_t\|y_{t+1}) \le \frac{1}{2}\eta^2 \operatorname{Tr}(x_t\nabla_t^2) \le \frac{\eta^2}{2}\|x_t\|_1\|\nabla_t\|^2 \le \frac{\eta^2 G_R^2}{2}.$$

□

Summing over t and telescoping and observing any eigenvalue is lower bounded by $\frac{1}{2^n T}$, we have

Regret

$$\begin{aligned}&\le \frac{1}{\eta}(B_R(\varphi_1\|x_1) - B_R(\varphi_T\|x_T)) \\ &\quad + \frac{1}{\eta}\sum_{t=1}^{T-1} |B_R(\varphi_t\|x_{t+1}) - B_R(\varphi_{t+1}\|x_{t+1})| + \frac{1}{\eta}\sum_{t=1}^{T-1} B_R(x_t\|y_{t+1}) \\ &\le \frac{2(D_R^2 + 2n + \log T)}{\eta} + \frac{\eta T G_R^2}{2} + \frac{1}{\eta}\sum_{t=1}^{T-1} |B_R(\varphi_t\|x_{t+1}) - B_R(\varphi_{t+1}\|x_{t+1})| \\ &\le \frac{2(D_R^2 + 2n + \log T)}{\eta} + \frac{\eta T G_R^2}{2} + \frac{1}{\eta}\sum_{t=1}^{T-1} |R(\varphi_t) - R(\varphi_{t+1}) - \nabla R(x_{t+1}) \bullet (\varphi_t - \varphi_{t+1})| \\ &\le \frac{2(D_R^2 + 2n + \log T)}{\eta} + \frac{\eta T G_R^2}{2} \\ &\quad + \frac{1}{\eta}\sum_{t=1}^{T-1} \|\nabla R(x_{t+1})\|\|(\varphi_t - \varphi_{t+1})\|_1 + \frac{1}{\eta}\sum_{t=1}^{T-1} |R(\varphi_t) - R(\varphi_{t+1})|. \end{aligned} \tag{12}$$

From [5] we already know $D_R = O(\sqrt{n})$ and $G_R = O(L)$ due to the use of the von Neumann entropy. We now compute an upper bound on $\|\nabla R(x_{t+1})\|$, suppose $x_{t+1} = VQV^\dagger$ is the eigendecomposition of x_{t+1}:

$$\|\nabla R(x_{t+1})\| = \|I + \log(x_{t+1})\| = \|I + V\log(Q)V^\dagger\| \leq n\log(2) + \log(T).$$

The inequality is due to the fact that for $x_{t+1} \in \mathcal{K}$, the eigenvalues of x_{t+1} are in $[\frac{1}{2^n T}, 1]$.

For the last term in the regret bound, the Fannes–Audenaert inequality [22] states that

$$|R(\varphi_t) - R(\varphi_{t+1})| \leq \frac{n}{2}\|\varphi_t - \varphi_{t+1}\|_1 + H(\frac{1}{2}\|\varphi_t - \varphi_{t+1}\|_1),$$

where $H(p) = -(p\log(p) + (1-p)\log(1-p))$ is the Shannon entropy.

We try to bound $H(p)$ by a linear function $g(p) = \lambda p + c$. Formally we have the following lemma:

Lemma 5. *For any $\lambda > 0$, we have for $p \in (0,1)$,*

$$H(p) \leq \lambda p + \log\frac{e^\lambda + 1}{e^\lambda}.$$

Proof. Take the derivative of $g(p) - H(p)$, we have that the minimal happens when $\lambda = \log\frac{1-p}{p}$ which is $p = \frac{1}{e^\lambda+1}$, and it's easy to verify that the minimal value of $g(p) - H(p)$ equals

$$\frac{\lambda}{e^\lambda+1} + c + \frac{1}{e^\lambda+1}\log\frac{1}{e^\lambda+1} + \frac{e^\lambda}{e^\lambda+1}\log\frac{e^\lambda}{e^\lambda+1}$$

by taking $c = \log\frac{e^\lambda+1}{e^\lambda}$, the minimal value equals zero and therefore $g(p)$ is an upper bound of $H(p)$. □

Choosing $\lambda = \log T$, we can upper bound the regret by

$$\textbf{Regret} \leq \frac{2(D_R^2 + 2n + \log T)}{\eta} + \frac{\eta T G_R^2}{2} + \frac{(n+\log T)\mathcal{P}}{\eta} + \frac{n\mathcal{P} + \log T\mathcal{P} + 1}{2\eta}. \tag{13}$$

Choosing $\eta = \sqrt{\frac{\mathcal{P}(n+\log T)}{TL^2}}$, the regret is upper bounded by $O(L\sqrt{T(n+\log T)\mathcal{P}})$.

Now we have proven the following dynamic regret bound:

$$\text{Dynamic Regret} = \sum_{t=1}^T \ell_t(x_t) - \min_{\varphi_t \in \mathcal{K}} \sum_{t=1}^T \ell_t(\varphi_t) = O(L\sqrt{T(n+\log T)\mathcal{P}}).$$

Notice that

$$\min_{\varphi_t \in \mathcal{K}} \sum_{t=1}^T \ell_t(\varphi_t) \leq \min_{\varphi_t \in C_n} \sum_{t=1}^T \ell_t(\varphi_t) + O(L),$$

we conclude that

$$\text{Dynamic Regret} = \sum_{t=1}^T \ell_t(x_t) - \min_{\varphi_t \in C_n} \sum_{t=1}^T \ell_t(\varphi_t) = O(L\sqrt{T(n+\log T)\mathcal{P}}).$$

□

For the rest of the section, we reason about choosing the optimal η, which depends on the actual path length that we do not know in advance. Algorithm 1 tackles this problem by constructing a set S of candidate learning rates, say $S = \{\frac{1}{4}, \frac{1}{8}, ...\}$, then initiate $|S|$ number of Algorithm 2 with different $\eta \in S$, and run weighted majority algorithm on top of these experts.

The regret of such a two-layer algorithm can be decomposed as a sum of the regret of the base expert and the meta MW algorithm [19]. Since the path length $\mathcal{P}$ is upper bounded by $O(T)$, we need only $|S| = O(\log T)$ base experts, and the additional regret induced by the weighted majority algorithm is only $O(\sqrt{T \log \log T})$. This regret is dominated by the dynamic regret of the best base expert, and thus is negligible in the final bound. Meanwhile, the optimal η is guaranteed to be either a small constant which is already good, or lie between two candidate learning rates so that one of them is near-optimal.

More concretely, the best expert has regret upper bounded by $O(L\sqrt{T(n + \log T)\mathcal{P}}) = \tilde{O}(L\sqrt{Tn\mathcal{P}})$ due to the existence of an optimal $\eta \in S$ by Lemma 3. The regret of the external MW algorithm in Algorithm 1 is upper bounded by $O(\sqrt{T \log(|S|)}) = \tilde{O}(\sqrt{T})$, a well-known regret bound of MW whose proof we omit here (see 27, Section 1.3 for example). Summing up, we get the final result $O(L\sqrt{T(n + \log(T))\mathcal{P}})$.

B Proof of Dynamic Regret Bounds Given Dynamical Models

B.1 Proof of Corollary 2

Algorithm 4: Dynamic Regret for Quantum Tomography Given Candidate Quantum Channels

Input: a candidate set of quantum channels $\{\Phi^{(j)}\}$,$1 \le j \le M$, a candidate set of η, $S = \{2^{-k-1} \mid 1 \le k \le \log T\}$, constant α.
Initialize $\log(T)M$ experts $E_{11}, ..., E_{\log(T)M}$, where E_{kj} is an instance of Algorithm 5 with $\eta = 2^{-k-1}$ and fixed quantum channel $\Phi^{(j)}$.
Set initial weights $w_1(k, j) = \frac{1}{\log(T)M}$.
for $t = 1, \dots, T$ **do**
    Predict $x_t = \frac{\sum_{k,j} w_t(k,j) x_t(k,j)}{\sum_{k,j} w_t(k,j)}$, where $x_t(k, j)$ is the output of expert E_{kj} at time t.
    Observe loss function $\ell_t(\cdot)$.
    Update the weights as

$$w_{t+1}(k, j) = w_t(k, j) e^{-\alpha \ell_t(x_t(k,j))}.$$

    Send gradients $\nabla \ell_t(x_t(k, j))$ to each expert E_{kj}.
end for

Similar to Appendix A, we also need to mix x with $\frac{I}{T2^n}$: $\tilde{x} = (1 - \frac{1}{T})x + \frac{I}{T2^n}$, for both x_t and φ_t. Notice that for any quantum channel Φ, we have $\Phi(\tilde{x}) = (1 - \frac{1}{T})\Phi(x) + \frac{I}{T2^n} = \widetilde{\Phi(x)}$, so we can use the same dynamical model in the new domain $\mathcal{K} = (1 - \frac{1}{T})C_n + \frac{1}{T2^n} I$ and the path length of the mixed comparators $\{\widetilde{\varphi_t}\}$ may only decrease. Due to the L-Lipschitzness of loss functions, the regret is only hurt by $O(LT\frac{1}{T}) = O(L)$.

We first prove a lemma which gives a dynamic regret bound of Algorithm 5 when the learning rate η is chosen optimally.

Algorithm 5: DMD for Quantum Tomography Given Fixed Quantum Channel

Input: domain $\mathcal{K} = (1 - \frac{1}{T})C_n + \frac{1}{T2^n}I$, step size $\eta < \frac{1}{2L}$, quantum channel $\Phi\colon C_n \mapsto C_n$.
Define $R(x) = \sum_{k=1}^{2^n} \lambda_k(x) \log \lambda_k(x)$, $\nabla R(x) := I + \log(x)$, and let B_R denote the Bregman divergence defined by R.
Set $x_1 = 2^{-n}I$, and y_1 to satisfy $\nabla R(y_1) = \mathbf{0}$.
for $t = 1, \ldots, T$ **do**
    Predict x_t and receive loss $\ell_t(\mathrm{Tr}(E_t x))$.
    Define $\nabla_t = \ell_t'(\mathrm{Tr}(E_t x_t))E_t$, where $\ell_t'(y)$ is a subderivative of ℓ_t with respect to y.
    Update y_{t+1} such that $\nabla R(y_{t+1}) = \nabla R(x_t) - \eta \nabla_t$.
    Update $\hat{x}_{t+1} = \mathrm{argmin}_{x \in \mathcal{K}} B_R(x || y_{t+1})$.
    Update $x_{t+1} = \Phi(\hat{x}_{t+1})$.
end for

Lemma 6. *Under the assumptions that the path length $\mathcal{P} = \sum_{t=1}^{T-1} \|\Phi(\varphi_t) - \varphi_{t+1}\|_1 \geq 1$ where $\|\cdot\|_1$ is the nuclear norm and $\Phi : C_n \mapsto C_n$ is a quantum channel, the dynamic regret of Algorithm 5 is upper bounded by $O(L\sqrt{T(n + \log T)\mathcal{P}})$, if we choose $\eta = \sqrt{\frac{\mathcal{P}(n+\log T)}{TL^2}}$.*

Proof. In general, quantum channels can be seen as completely positive trace-preserving (CPTP) maps on the space of density operators. Ref. [42] gave the following result:

Lemma 7 (Contraction of CPTP maps with respective to relative entropy 42)**.** *Let $T(\mathcal{H})$ be the operators in a Hilbert space $\mathcal{H}$ with finite trace and $T_+(\mathcal{H})$ be the positive elements in $T(\mathcal{H})$. If Φ is a completely positive trace-preserving map of $T(\mathcal{H})$ into itself, then for all $A, B \in T_+(\mathcal{H})$*

$$S(\Phi(A)\|\Phi(B)) \leq S(A\|B),$$

where $S(A\|B)$ is the relative entropy between A and B.

Since density operators on finite Hilbert space is positive and has finite trace, and $B_R(x\|y) = S(x\|y)$ when R is the von Neumann entropy, we can derive that if $\Phi(\cdot)\colon C_n \to C_n$ is a quantum channel on C_n,

$$\forall x, y \in C_n, B_R(\Phi(x)\|\Phi(y)) \leq B_R(x\|y). \tag{14}$$

Writing $\ell_t(x) = \ell_t(\mathrm{Tr}(E_t x))$, similar to Eq. (10), we can prove that

$$\begin{aligned}
&\ell_t(x_t) - \ell_t(\varphi_t) \\
\leq &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_t\|y_{t+1}) + B_R(x_t\|y_{t+1})) \\
\leq &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_t\|\hat{x}_{t+1}) + B_R(x_t\|y_{t+1})) \\
= &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_{t+1}\|x_{t+1}) + B_R(\varphi_{t+1}\|x_{t+1}) - B_R(\varphi_t\|\hat{x}_{t+1}) + B_R(x_t\|y_{t+1})) \\
\leq &\frac{1}{\eta}(B_R(\varphi_t\|x_t) - B_R(\varphi_{t+1}\|x_{t+1}) + B_R(\varphi_{t+1}\|x_{t+1}) - B_R(\Phi(\varphi_t)\|x_{t+1}) + B_R(x_t\|y_{t+1})),
\end{aligned}$$

where the fourth inequality comes from Eq. (14).

By telescoping sums, we get

$$\begin{aligned}
&\textbf{Regret}\\
\leq&\frac{1}{\eta}(B_R(\varphi_1\|x_1)-B_R(\varphi_T\|x_T))\\
&+\frac{1}{\eta}\sum_{t=1}^{T-1}|B_R(\Phi(\varphi_t)\|x_{t+1})-B_R(\varphi_{t+1}\|x_{t+1})|+\frac{1}{\eta}\sum_{t=1}^{T-1}B_R(x_t\|y_{t+1})\\
\leq&\frac{1}{\eta}B_R(\varphi_1\|x_1)+\frac{\eta TG_R^2}{2}+\frac{1}{\eta}\sum_{t=1}^{T-1}|B_R(\Phi(\varphi_t)\|x_{t+1})-B_R(\varphi_{t+1}\|x_{t+1})|.
\end{aligned}$$

Notice that this inequality is the same as Eq. (12) replacing φ_t by $\Phi(\varphi_t)$, so we can follow the same proof of Eq. (13) and obtain a variant regret bound

$$\textbf{Regret}\leq\frac{D_R^2}{\eta}+\frac{\eta TG_R^2}{2}+\frac{(n+\log T)\mathcal{P}}{\eta}+\frac{n\mathcal{P}+\log T\mathcal{P}+1}{2\eta},$$

where $\mathcal{P}=\sum_{t=1}^{T-1}\|\varphi_{t+1}-\Phi(\varphi_t)\|_1$.

Choosing $\eta=\sqrt{\frac{\mathcal{P}(n+\log T)}{TL^2}}$, the regret is upper bounded by $O(L\sqrt{T(n+\log T)\mathcal{P}})$. □

Proof of Corollary 2. We will prove that Algorithm 4 can achieve the dynamic regret bound in this corollary and in the following proof we will use the notation in the description of Algorithm 4.

In Algorithm 4, we run weighted majority algorithm on $M\log(T)$ experts, each of which runs Algorithm 5 with quantum channel $\Phi^{(j)}$ and step size in $\{\frac{1}{4},\frac{1}{8},\ldots\}$.

From Lemma 6, we can conclude that the minimum dynamic regret bound of all $M\log(T)$ experts is

$$O(L\sqrt{T(n+\log(T))\mathcal{P}'}),$$

where $\mathcal{P}'=\min_{\Phi^{(j)}\in\{\Phi^{(1)},\ldots,\Phi^{(M)}\}}\sum_{t=1}^{T-1}\|\varphi_{t+1}-\Phi^{(j)}(\varphi_t)\|_1$.

In Lemma 1 of the arXiv version of [28], they give a regret bound of the weighted majority algorithm:

$$\sum_{t=1}^{T}\ell_t(x_t)\leq\sum_{t=1}^{T}\ell_t(x_t(k,j))+\frac{c\sqrt{2T\ln\left(\frac{1}{w_1(k,j)}\right)}}{4},$$

for any $1\leq k\leq\log(T),1\leq j\leq M$ choosing $\alpha=\sqrt{\frac{8}{Tc^2\ln\left(\frac{1}{w_1(k,j)}\right)}}$.

Let k,j be the optimal one of all $M\log(T)$ experts, we have

$$\begin{aligned}
\sum_{t=1}^{T}\ell_t(x_t)\leq&\sum_{t=1}^{T}\ell_t(\varphi_t)+c_1L\sqrt{T(n+\log(T))\mathcal{P}'}+\frac{c\sqrt{T\ln(\log(T)M)}}{4}\\
=&\sum_{t=1}^{T}\ell_t(\varphi_t)+O(L\sqrt{T(n+\log(T))\mathcal{P}'}+\sqrt{T\log(M\log(T))}),
\end{aligned}$$

which concludes our result. □

B.2 Proof of Corollary 3

In general, the possible dynamical models can be a continuous set, and we can take it to be the set of all local quantum channels for example. A quantum channel on n qubits is said to be l-local if it can be written as a product channel $\Phi = \Phi_1 \otimes I$, where Φ_1 only influences l qubits.

Lemma 8 (Stinespring's dilation theorem 43)**.** *Let $\Phi : C_n \to C_n$ be a completely positive and trace-preserving map on the space of density operators of a Hilbert space $\mathcal{H}_s$. Then there exists a Hilbert space $\mathcal{H}_e$ with $\dim \mathcal{H}_e = (\dim \mathcal{H}_s)^2$ and a unitary operation U on $\mathcal{H}_s \otimes \mathcal{H}_e$ such that*

$$\Phi(\rho) = \mathrm{Tr}_e(U(\rho \otimes |0\rangle\langle 0|_e)U^\dagger)$$

for all $\rho \in C_n$, where Tr_e denotes the partial trace on $\mathcal{H}_e$.

By Stinespring's dilation theorem, we can represent the quantum channel on a 2^l dimension Hilbert space $\mathcal{H}_s$ by partial trace of a pure quantum channel on a 2^{3l} dimension Hilbert space $\mathcal{H}_s \otimes \mathcal{H}_e$, so we only need to find ϵ-net of unitary operators.

Lemma 9 (ϵ-net of U_l 44, Lemma 4.4)**.** *Let U_l be the set of unitaries on Hilbert space $\mathcal{H}$ with $\dim \mathcal{H} = 2^l$. For $\delta > 0$, there is a subset $U_{l,\delta}$ of U_l with*

$$|U_{l,\delta}| \leq \left(\frac{2}{\delta}\right)^{2^{4l}}$$

such that for any $U \in U_l$, there exists $V \in U_{l,\delta}$ with $\|U - V\|_2 \leq \delta$.

From Lemma 8 and Lemma 9, we can construct an ϵ-net of general quantum channels:

Lemma 10 (ϵ-net of quantum channels)**.** *Let S_l be the set of all quantum channels on density operators on Hilbert space $\mathcal{H}$ with $\dim \mathcal{H} = 2^l$. For any $1 > \delta > 0$, there is a subset $S_{l,\delta}$ of S_l with*

$$|S_{l,\delta}| \leq \left(\frac{2}{\delta}\right)^{2^{12l}},$$

such that for any $\Phi \in S_l$, there exists $\tilde{\Phi} \in S_{l,\delta}$ such that $\sup_{x \in C_l} \|\Phi(x) - \tilde{\Phi}(x)\|_1 \leq 3\delta$.

Proof. Let $\mathcal{H}_e$ be a Hilbert space with $\dim(\mathcal{H}_e) = 2^{2l}$ and $\mathcal{H}_a = \mathcal{H} \otimes \mathcal{H}_e$. Let $U_{3l,\delta}$ be the ϵ-net of unitaries on $\mathcal{H}_a$ in Lemma 9.

Then, we can construct $S_{l,\delta} = \{\Phi \mid \exists U \in U_{3l,\delta}, \forall x \in C_l, \Phi(x) = \mathrm{Tr}_e(U(x \otimes |0\rangle\langle 0|_e)U^\dagger\}$.

For an arbitrary $\Phi' \in S_l$, from Lemma 8, we can show that there exists a Hilbert space $\mathcal{H}_c$ with $\dim \mathcal{H}_c = 2^{2l}$ and a unitary operation U' on $\mathcal{H} \otimes \mathcal{H}_c$ such that

$$\Phi(x) = \mathrm{Tr}_c(U'(x \otimes |0\rangle\langle 0|_c)U'^\dagger)$$

for any $x \in C_l$.

Notice that $\mathcal{H} \otimes \mathcal{H}_e$ and $\mathcal{H} \otimes \mathcal{H}_c$ are isomorphic since they are Hilbert space with the same dimension, so there exists $V' \in U_{3l,\delta}$ such that $\|U' - V'\|_2 \leq \delta$.

From the definition of $S_{l,\delta}$, there exists $\tilde{\Phi} \in S_{l,\delta}$ such that

$$\tilde{\Phi}(x) = \mathrm{Tr}_e(V'(x \otimes |0\rangle\langle 0|_e)V'^\dagger)$$

for any $x \in C_l$.

Then, we can show that for any $x \in C_l$

$$\begin{aligned}
&\|\Phi(x) - \tilde{\Phi}(x)\|_1 \\
=&\|\operatorname{Tr}_e(U'(x \otimes |0\rangle\langle 0|_e)U'^\dagger - V'(x \otimes |0\rangle\langle 0|_e)V'^\dagger)\|_1 \\
\leq&\|U'(x \otimes |0\rangle\langle 0|_e)U'^\dagger - V'(x \otimes |0\rangle\langle 0|_e)V'^\dagger\|_1 \\
\leq&2\|(U'-V')(x \otimes |0\rangle\langle 0|_e)V'^\dagger\|_1 + \|(U'-V')(x \otimes |0\rangle\langle 0|_e)(U'-V')^\dagger\|_1 \\
\leq&2\|U'-V'\|_2\|(x \otimes |0\rangle\langle 0|_e)V'^\dagger\|_1 + \|U'-V'\|_2^2\|x \otimes |0\rangle\langle 0|_e\|_1 \leq 2\delta + \delta^2 \leq 3\delta,
\end{aligned}$$

where in the first equality, we let Φ be a quantum channel on the density operators on $\mathcal{H} \otimes \mathcal{H}_e$ since $\mathcal{H}_s$ and $\mathcal{H}_e$ are isomorphic.

The size of $S_{l,\delta}$ is the same as $|U_{3l,\delta}|$ which is $\left(\frac{2}{\delta}\right)^{2^{12l}}$. □

With Lemma 10, we can construct an ϵ-net of l-local quantum channels.

Lemma 11 (ϵ-net of l-local quantum channels)**.** *Let $S_{n,l}$ be the set of all l-local quantum channels on n qubits. For any $1 > \delta > 0$, there is a subset $S_{n,l,\delta}$ of $S_{n,l}$ with*

$$|S_{n,l,\delta}| \leq \binom{n}{l}\left(\frac{6}{\delta}\right)^{2^{12l}},$$

such that for any $\Phi \in S_{n,l}$, there exists $\tilde{\Phi} \in S_{n,l,\delta}$ such that $\sup_{x \in C_n} \|\Phi(x) - \tilde{\Phi}(x)\|_1 \leq \delta$.

Proof. From Lemma 10, we can construct a collection of sets $\{W_{i_1,\ldots,i_l} \mid 1 \leq i_1 < \cdots < i_l \leq n\}$. Each $W_{i_1,\ldots,i_l}$ is a set of quantum channels on C_n and contains extension of channels in $S_{l,\frac{\delta}{3}}$ in Lemma 10 acting on qubits $i_1, \ldots, i_l$ and all operators in it perform identity on the other $n-l$ qubits, so $|W_{i_1,\ldots,i_l}| = \left(\frac{6}{\delta}\right)^{2^{12l}}$.

Let $S_{n,l,\delta}$ be the union of all W in $\{W_{i_1,\ldots,i_l} \mid 1 \leq i_1 < \cdots < i_l \leq n\}$, then we can conclude for any quantum channel $\Phi \in S_{n,l}$, there exists $\tilde{\Phi}(x) \in S_{n,l,\delta}$ with $\sup_{x \in C_n} \|\Phi(x) - \tilde{\Phi}(x)\|_1 \leq \delta$ and $|S_{n,l,\delta}| \leq \sum_{i_1,\ldots,i_l} |W_{i_1,\ldots,i_l}| \leq \binom{n}{l}\left(\frac{6}{\delta}\right)^{2^{12l}}$. □

Proof of Corollary 3. Let

$$\mathcal{P}' = \min_{\Phi \in S_{n,l}} \sum_{t=1}^{T-1} \|x_{t+1} - \Phi(x_t)\|_1, \tag{15}$$

where $S_{n,l}$ is the set of all l-local quantum channels on n qubits.

For any $\Phi \in S_{n,l}$, there exists $\tilde{\Phi}$ such that $\sup_{x \in C_n} \|\Phi(x) - \tilde{\Phi}(x)\|_1 \leq \delta$, then we have

$$\|x_{t+1} - \Phi(x_t)\|_1 \geq \|x_{t+1} - \tilde{\Phi}(x_t)\|_1 - \|\Phi(x) - \tilde{\Phi}(x)\|_1 \geq \|x_{t+1} - \tilde{\Phi}(x_t)\|_1 - \delta,$$

for any $1 \leq t \leq T$. Thus, we have

$$\min_{\Phi \in S_{n,l}} \sum_{t=1}^{T-1} \|x_{t+1} - \Phi(x_t)\|_1 \geq \min_{\Phi \in S_{n,l,\delta}} \sum_{t=1}^{T-1} \|x_{t+1} - \Phi(x_t)\|_1 - T\delta.$$

From Corollary 2, we can show that the regret of Algorithm 4 with the candidate set of quantum channels to be the $S_{n,l,\delta}$ in Lemma 11 satisfies

Regret

$$\leq c_1 L\sqrt{T(n+\log(T))\min_{\Phi\in S_{n,l,\delta}}\sum_{t=1}^{T-1}\|x_{t+1}-\Phi(x_t)\|_1}+c_2\sqrt{T\log\log(T)+T\log\left(\binom{n}{l}\left(\frac{6}{\delta}\right)^{2^{12l}}\right)}$$

$$\leq c_1 L\sqrt{T(n+\log(T))(\min_{\Phi\in S_{n,l}}\sum_{t=1}^{T-1}\|x_{t+1}-\Phi(x_t)\|_1+T\delta)}+c_2\sqrt{T\log\log(T)+T\log\left(\binom{n}{l}\left(\frac{6}{\delta}\right)^{2^{12l}}\right)}$$

$$= c_1 L\sqrt{T(n+\log(T))(\mathcal{P}'+3T\delta)}+c_2\sqrt{T\log\log(T)+T\left(2^{12l}\log\left(\frac{6}{\delta}\right)+l\log(n)\right)}.$$

Let $\delta=\frac{1}{T}$, we can achieve a regret of $\tilde{O}(L\sqrt{nT\mathcal{P}'}+2^{6l}\sqrt{T})$ when $\mathcal{P}'\geq 1$. □

C Other Proofs

C.1 K-outcome Measurement Cases

We can generalize the two-outcome measurements in our results to K-outcome measurements. The only issue of replacing two-outcome measurements with K-outcome measurements is that the output of the K-outcome measurement is a probability distribution rather than a scalar, so we need a new loss function to measure the regret. Let $\mathbf{p}_t=(\mathrm{Tr}(E_{t,1}x_t),\ldots,\mathrm{Tr}(E_{t,K}x_t))$ and $\mathbf{b}_t=(b_{t,i})_{i=1}^K$ be the probability distribution corresponding to the output of the measurement actting on the ground truth state ρ_t. In this case, we can define the loss function at time t to be

$$\ell_t(x_t)=d_{\mathrm{TV}}(\mathbf{p_t},\mathbf{b_t})=\frac{1}{2}\sum_{i=1}^{K}\left|\mathrm{Tr}\left(E_{t,i}x_t\right)-b_{t,i}\right|. \tag{16}$$

This is a convex function so the proof in previous sections still holds and we only need to determine G_R which is the upper bound on $\|\nabla_t\|$.

First, we prove the following inequality on the spectral norm of linear combination of E_i. For any K-outcome measurement $\mathcal{M}$ described by $\{E_1,\ldots,E_K\}$ and K real numbers $a_1,\ldots,a_K$ such that $|a_i|\leq L$ for all $i\in[K]$, we have

$$\begin{aligned}\|\sum_{i=1}^{K}a_iE_i\| &= \max_{v\in\mathbb{C}^n,\|v\|_2=1}v^\dagger(\sum_{i=1}^{K}a_iE_i)v\leq\max_{v\in\mathbb{C}^n,\|v\|_2=1}\sum_{i=1}^{K}|a_i|v^\dagger E_iv\\ &\leq\max_{v\in\mathbb{C}^n,\|v\|_2=1}Lv^\dagger(\sum_{i=1}^{K}E_i)v=L,\end{aligned} \tag{17}$$

where we use spectral norm in the first line and the first inequlity comes from $0\preceq E_i$. Then we have

$$\|\nabla_t\|=\|\sum_{i=1}^{K}\frac{1}{2}\mathrm{sgn}(\mathrm{Tr}(E_{t,i}x_t)-b_{t,i})E_{t,i}\|\leq\frac{1}{2}, \tag{18}$$

where $\mathrm{sgn}(x)$ is the sign function.

Therefore, $G_R=\frac{1}{2}$ is a valid parameter in this case. Since we have $G_R=L$ in previous two-outcome cases, we can simply repalce all L with $\frac{1}{2}$ in our results in two-outcome cases to obatin an algorithm in K-outcome cases with loss function in Eq. (16).

C.2 Proof of Corollary 1

Proof. To achieve the mistake bound, we run $\mathcal{A}$ in a subset of the iterations. Specifically, we only update x_t if $\ell_t(\mathrm{Tr}(E_t x_t)) > \frac{2}{3}\epsilon$. Since $\ell_t(\mathrm{Tr}(E_t x_t)) > \frac{2}{3}\epsilon$ whenever $|\mathrm{Tr}(E_t x_t) - \mathrm{Tr}(E_t \rho_t)| > \epsilon$, the number of the mistakes $\mathcal{A}$ makes is at most the number of the updates $\mathcal{A}$ makes.

Let the sequence of t when $\mathcal{A}$ updates x_t be $\{t_i\}_{i=1}^T$. Since $\{\rho_{t_i}\}$ is a valid comparator sequence (sub-sequence of a k-shift sequence shifts at most k times) with $\sum_{i=1}^T \ell_{t_i}(\mathrm{Tr}(E_{t_i}\rho_{t_i})) \leq \frac{1}{3}\epsilon T$, the number of updates T satisfies the inequality $\frac{2}{3}\epsilon T \leq \frac{1}{3}\epsilon T + \tilde{O}(L\sqrt{Tn\mathcal{P}})$, where $\tilde{O}(L\sqrt{Tn\mathcal{P}})$ comes from the fact that the algorithm's regret compared to the best expert is upper bounded by $\tilde{O}(L\sqrt{Tn\mathcal{P}})$. Thus, it implies $T = \tilde{O}(\frac{L^2 n\mathcal{P}}{\epsilon^2})$, which completes the proof. □

C.3 Proof of Lemma 1

Proof. The CBCE algorithm in [20] is a meta-algorithm which uses any online convex optimization algorithm $\mathcal{B}$ as a black-box algorithm and obtain a strongly adaptive regret with an $O(\sqrt{\tau \log(T)})$ overhead for any interval length τ. Its guarantee is formalized in the following lemma.

Lemma 12 (SA-Regret of CBCE$\langle\mathcal{B}\rangle$, 20, Theorem 2)**.** *Assume the loss function $\ell_t(x)$, $x \in \mathcal{K}$ is convex and maps to $[0,1], \forall t \in \mathbb{N}$ and that the black-box algorithm $\mathcal{B}$ has static regret $R_T^{\mathcal{B}}$ bounded by cT^α, where $\alpha \in (0,1)$. Let $I = [I_1, I_2]$. The SA-Regret of CBCE with black-box $\mathcal{B}$ on the interval I is bounded as follows:*

$$\begin{aligned} R^{\mathrm{CBCE}\langle\mathcal{B}\rangle}(I) &= \min_{\varphi\in\mathcal{K}} \sum_{t\in I} \left(\ell_t\left(\mathbf{x}_t^{\mathcal{M}\langle\mathcal{B}\rangle}\right) - \ell_t(\varphi)\right) \leq \frac{4}{2^\alpha - 1} c|I|^\alpha + 8\sqrt{|I|\left(7\log\left(I_2\right) + 5\right)} \\ &= O\left(c|I|^\alpha + \sqrt{|I|\log(I_2)}\right). \end{aligned}$$

where $\mathcal{M}\langle\mathcal{B}\rangle$ denotes the meta-algorithm. Note that this bound can be obtained with any bounded convex function sequence ℓ_t.

For the choice of black-box algorithm, the RFTL based algorithm we used as the black-box function in [5] achieves an $O(\sqrt{Tn})$ bound on the static regret when ℓ_t are ℓ_1 or ℓ_2 loss.

Lemma 13 (5, Theorem 2)**.** *Let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements on an n-qubit state presented to the learner, and suppose ℓ_t is convex and L-Lipschitz. Then there is an explicit learning strategy that guarantees regret $R_T = O(L\sqrt{Tn})$ for all T. Specifically, when applied to ℓ_1 loss and ℓ_2 loss, the RFTL algorithm achieves regret $O(\sqrt{Tn})$ for both.*

Note that both ℓ_1 and ℓ_2 norm versions of our loss function ℓ_t are 2-Lipschitz and bounded. Therefore, we can directly obtain an algorithm with strongly adaptive regret bound by combining Lemma 12 and Lemma 13 with $\alpha = 1/2$. □

C.4 Proof of Theorem 2

Proof. Given any $1 = t_1 \leq t_2 \leq \cdots \leq t_{k+1} = T + 1$ and $\varphi_1, \varphi_2, \ldots, \varphi_k \in C_n$, $\mathcal{A}$ guarantees that

$$\begin{aligned}
&\sum_{t=1}^{T} \ell_t(E_t x_t^{\mathcal{A}}) - \sum_{j=1}^{k} \sum_{t=t_j}^{t_{j+1}-1} \ell_t(E_t \varphi_j) \\
\leq &\sum_{j=1}^{k} cL\sqrt{n(t_{j+1} - t_j)\log(T)} \\
\leq &cL\sqrt{knT\log(T)},
\end{aligned}$$

where c is some constant. In the second line we use the result of Lemma 1, and in the third line we use the Cauchy–Schwarz inequality. □

C.5 Proof of Corollary 4

Proof. To achieve the mistake bound, we run $\mathcal{A}$ in a subset of the iterations. Specifically, we only update x_t if $\ell_t(\mathrm{Tr}(E_t x_t)) > \frac{2}{3}\epsilon$. Since $\ell_t(\mathrm{Tr}(E_t x_t)) > \frac{2}{3}\epsilon$ whenever $|\mathrm{Tr}\,(E_t x_t) - \mathrm{Tr}\,(E_t \rho_t)| > \epsilon$, the number of the mistakes $\mathcal{A}$ makes is at most the number of the updates $\mathcal{A}$ makes.

Let the sequence of t when $\mathcal{A}$ updates x_t be $\{t_i\}_{i=1}^{T}$. Since $\{\rho_{t_i}\}$ is a valid comparator sequence (sub-sequence of a k-shift sequence shifts at most k times) with $\sum_{i=1}^{T} \ell_{t_i}(\mathrm{Tr}(E_{t_i}\rho_{t_i})) \leq \frac{1}{3}\epsilon T$, from Theorem 2, we can conclude that the number of updates T satisfies the inequality $\frac{2}{3}\epsilon T \leq \frac{1}{3}\epsilon T + c\sqrt{knT\log(T)}$. Thus, it implies $T = O(\frac{kn}{\epsilon^2}\log^2(\frac{kn}{\epsilon^2}))$, which completes the proof. □