

# Online Self-Concordant and Relatively Smooth Minimization, With Applications to Online Portfolio Selection and Learning Quantum States

Chung-En Tsai<sup>1</sup>, Hao-Chung Cheng<sup>2,3,4,5,6</sup>, and Yen-Huan Li<sup>1,3,5</sup>

<sup>1</sup>Department of Computer Science and Information Engineering,  
National Taiwan University

<sup>2</sup>Department of Electrical Engineering and Graduate Institute of Communication  
Engineering, National Taiwan University

<sup>3</sup>Department of Mathematics, National Taiwan University

<sup>4</sup>Hon Hai (Foxconn) Quantum Computing Centre

<sup>5</sup>Center for Quantum Science and Engineering,  
National Taiwan University

<sup>6</sup>Physics Division, National Center for Theoretical Sciences,  
Taipei 10617, Taiwan

## Abstract

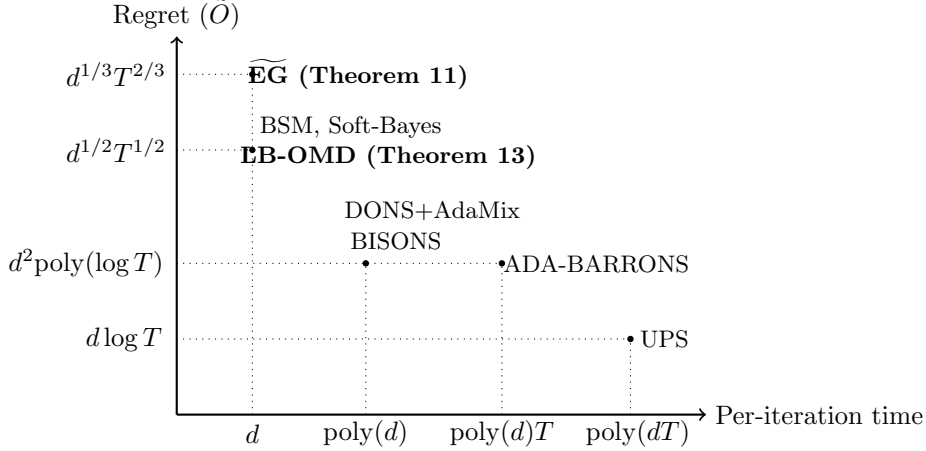
Consider an online convex optimization problem where the loss functions are self-concordant barriers, smooth relative to a convex function  $h$ , and *possibly non-Lipschitz*. We analyze the regret of online mirror descent with  $h$ . Then, based on the result, we prove the following in a *unified* manner. Denote by  $T$  the time horizon and  $d$  the parameter dimension.

1. For online portfolio selection, the regret of  $\widetilde{\text{EG}}$ , a variant of exponentiated gradient due to Helmbold et al. [1], is  $\tilde{O}(T^{2/3}d^{1/3})$  when  $T > 4d/\log d$ . This improves on the original  $\tilde{O}(T^{3/4}d^{1/2})$  regret bound for  $\widetilde{\text{EG}}$ .
2. For online portfolio selection, the regret of online mirror descent with the logarithmic barrier is  $\tilde{O}(\sqrt{Td})$ . The regret bound is the same as that of Soft-Bayes due to Orseau et al. [2] up to logarithmic terms.
3. For online learning quantum states with the logarithmic loss, the regret of online mirror descent with the log-determinant function is also  $\tilde{O}(\sqrt{Td})$ . Its per-iteration time is shorter than all existing algorithms we know.

## 1 Introduction

Online portfolio selection (OPS) is an online convex optimization problem more than three decades old. Unlike standard online convex optimization problems, the loss functions in OPS are neither Lipschitz nor smooth. Existing analyses of standard algorithms, such as online mirror descent (OMD) and follow the regularized leader (FTRL), are hence not directly applicable. The optimal regret in OPS is known to be  $O(d \log T)$ , where  $d$  denotes the parameter dimension and  $T$  the time horizon, and achieved by Universal Portfolio Selection (UPS) [3, 4]. The per-iteration time of UPS, however, is too high for the algorithm to be practical [5].  $O(d)$  is perhaps the lowest per-iteration time one can expect; this is actually achieved by  $\widetilde{\text{EG}}$  [1], the barrier subgradient method (BSM) [6], and Soft-Bayes [2]. Nevertheless, the regret bound for  $\widetilde{\text{EG}}$  is  $\tilde{O}(d^{1/2}T^{3/4})$ ; the regret bound for BSM and Soft-Bayes is  $\tilde{O}(\sqrt{Td})$ . Both are significantly higher than that of UPS in terms of the dependence on  $T$ . Recent researches on OPS focus on achieving logarithmic regret (in  $T$ ) with acceptable per-iteration time. ADA-BARRONS achieves  $O(d^2 \text{poly}(\log T))$  regret with  $O(\text{poly}(d)T)$  per-iteration time [7]. DONS with AdaMix [8] and BIONS [9] both possess the same  $O(d^2 \text{poly}(\log T))$  regret guarantee and reduce the per-iteration time to  $O(\text{poly}(d))$ . We summarize existing and our results in the form of an “efficiency-regret Pareto frontier” in Figure 1.

Figure 1: Current efficiency-regret Pareto frontier for OPS, assuming  $T \geq 4d/\log d$ . The figure is modified from the one by Zimmert et al. [9].



The generalization of OPS for the quantum setup (Section 3.3), interestingly, corresponds to online learning quantum states. This generalization can be understood either as online maximum-likelihood quantum state estimation [10] or a quantum version of probability forecasting with the logarithmic loss [9]. The quantum generalization is even more challenging: The dimension grows exponentially with the number of qubits (quantum bits). Scalability with respect to the dimension is hence critical.

To the best of our knowledge, there are only two algorithms for the quantum generalization. Both are obtained via generalizing existing OPS algorithms.

1. Q-Soft-Bayes [10] is a quantum generalization of Soft-Bayes. It inherits the  $O(\sqrt{Td \log d})$  regret guarantee of Soft-Bayes with  $\tilde{O}(d^3)$  per-iteration time.
2. Schrödinger’s-BISIONS [9] is a quantum generalization of BISIONS. It has  $O(d^3 \log^2 T)$  regret and  $\tilde{O}(d^6)$  per-iteration time<sup>1</sup>.

In OPS, unless the dimension is extremely high, one would prefer BISIONS to Soft-Bayes. In the quantum generalization, regarding the exponentially growing dimension, the situation differs. It turns out that Q-Soft-Bayes’ scalability with respect to the dimension makes it competitive.

In this paper, instead of pursuing a logarithmic regret with acceptable dimension scalability, we focus on sublinear-regret algorithms that scale with dimension. In particular, we prove the following.

1. For OPS, the regret of  $\widetilde{\text{EG}}$ , a variant of exponentiated gradient proposed by Helmbold et al. [1], is indeed  $\tilde{O}(d^{1/3}T^{2/3})$  when  $T > 4d/\log d$ . This improves upon the  $\tilde{O}(d^{1/2}T^{3/4})$  regret bound by Helmbold et al. [1].
2. For OPS, LB-OMD (OMD with the logarithmic barrier) yields  $\tilde{O}(\sqrt{dT})$  regret with  $\tilde{O}(d)$  per-iteration time when  $T > d$ . Both are the same as those of BSM and Soft-Bayes.
3. For the quantum generalization of OPS, Q-LB-OMD (OMD with the log-det function) yields  $\tilde{O}(\sqrt{dT})$  regret with  $\tilde{O}(d^3)$  per-iteration time when  $T > d$ . Both the regret bound and time cost are the same as those of Q-Soft-Bayes. Nevertheless, each iteration of Q-Soft-Bayes requires computing one matrix exponential and two matrix logarithms, whereas each iteration of Q-LB-OMD only requires computing one eigendecomposition; the time cost of all other operations in the two algorithms are negligible.

<sup>1</sup>Zimmert et al. [9] only claims a  $O(\text{poly}(d))$  per-iteration time, achieved by second-order convex optimization methods. We hence evaluate the per-iteration time as  $\tilde{O}((d^2)^3)$ .

The three results above are derived in a *unified* manner. We approach OPS and its quantum generalization via analyzing the regret of OMD for loss functions that are relatively smooth, *possibly non-Lipschitz*, self-concordant barriers. This is motivated by the fact that the loss functions in OPS are standard instances of self-concordant barriers [11] and of relatively smooth functions [12], simultaneously. For online convex optimization with such loss functions, we derive a regret bound for OMD. Then, we prove the three results mentioned above as simple consequences of this result, thereby providing a principled approach to achieving the performances of Soft-Bayes and BSM and slightly better performances than Q-Soft-Bayes.

Online portfolio selection and online learning quantum states are two special cases covered by our framework. In most relatively smooth convex optimization problems, the loss functions are also self-concordant barriers. Therefore, our framework also applies to most problems studied by Bauschke et al. [12], Lu et al. [13], and Eshraghi and Liang [14] for relatively smooth convex optimization, in their online convex optimization formulations.

**Notation** Let  $N \in \mathbb{N}$ . We write  $[N]$  for the set  $\{1, \dots, N\}$ . We write  $\mathbb{R}_+$  and  $\mathbb{R}_{++}$  for the sets of non-negative numbers and strictly positive numbers, respectively. We denote by  $\mathbb{S}^d$ ,  $\mathbb{S}_+^d$ , and  $\mathbb{S}_{++}^d$  the set of Hermitian matrices, Hermitian positive semi-definite matrices, and Hermitian positive definite matrices in  $\mathbb{C}^{d \times d}$ , respectively. For any vector  $v \in \mathbb{R}^d$ , we write  $v(i)$  for its  $i$ -th entry. Let  $f : \mathbb{R}^d \rightarrow [-\infty, \infty]$  be an extended-valued function. Its domain is denoted by  $\text{dom } f$ . The interior and relative interior of a set  $\mathcal{S}$  are denoted by  $\text{int } \mathcal{S}$  and  $\text{ri } \mathcal{S}$ , respectively. Let  $H \in \mathbb{S}^d$  with eigendecomposition  $H = \sum_i \lambda_i P_i$ , where  $\lambda_i$  are eigenvalues and  $P_i$  are projections. Let  $f$  be a function whose domain contains  $\{\lambda_i\}$ . Then,  $f(H) := \sum_i f(\lambda_i) P_i$ .

## 2 Related Work

Relevant literature on OPS have been reviewed in Section 1.

### 2.1 Online Learning on the Set of Quantum States

Warmuth and Kuzmin [15] and Arora and Kale [16] independently discovered OMD with the von Neumann entropy for online principal component analysis (PCA) and semi-definite programming, respectively. The loss functions they considered are linear. Koolen et al. [17] studied a quantum generalization of probability forecasting with the *matrix entropic loss*. Whereas the problem formulation looks similar to the quantum generalization of OPS, the specific matrix entropic loss admits simple solutions based on existing probability forecasting algorithms. Another problem closely related to the quantum generalization is online shadow tomography [18, 19]. Online shadow tomography corresponds to online convex optimization with the absolute loss on the set of quantum states. The absolute loss is Lipschitz and hence standard from the online convex optimization perspective. Yang et al. [20] studied online convex optimization on the set of quantum states with general Lipschitz losses. The most relevant to this paper is perhaps the work on bandit PCA by Kotłowski and Neu [21]. As in this paper, Kotłowski and Neu [21] also considered OMD with the log-det function and showed that its iteration rule, though lacking a closed-form expression, can be computed efficiently.

### 2.2 Self-Concordance

The notions of self-concordance and self-concordant barriers were originally proposed for analyzing and developing convex optimization algorithms [11]. Self-concordant barriers have been shown useful as regularizers in several online learning scenarios, such as bandit linear optimization [22] and optimistic online learning [23]; see also the monograph by Hazan [24]. In online learning, examples of self-concordant barriers as loss functions include probability forecasting [25, Chapter 9], OPS, and the quantum generalization of OPS. Except for the analysis by van Erven et al. [26], existing results do not explicitly use general properties of self-concordant barrier. Zhang et al. [27] studied the dynamic regret with losses that are self-concordant *and* strictly convex; it is easily checked that the strict convexity assumption is violated in OPS and its quantum generalization.

## 2.3 Relative Smoothness

The notion of relative smoothness was also originally proposed for analyzing and developing convex optimization algorithms, as a generalization of smoothness [12, 13]. Although smoothness is a standard assumption in convex optimization literature, this notion is relatively rarely exploited in online convex optimization. A few examples include the  $L^*$  bound [28, 29] and studies on the dynamic regret by, e.g., Jadbabaie et al. [30], Mokhtari et al. [31], Zhang et al. [27]. There seems to be only one paper on online convex optimization with relatively smooth losses [14]. The paper considers dynamic regret that in the worst case does not guarantee a sublinear regret; its requirement of a Lipschitz Bregman divergence is violated in LB-OMD.

## 2.4 Comparison with BSM and LB-FTRL

Our analysis exploits the properties of self-concordant barriers, which may look familiar to online learning experts. The use of self-concordant barriers in online learning was perhaps popularized by the “interior-point methods” proposed by Abernethy et al. [32]. For OPS, the analyses of BSM [6] and LB-FTRL [26] also exploit the properties of self-concordant barriers. The difference lies in where the barrier properties are used. In the analyses of the interior-point methods, BSM, and LB-FTRL, it is the regularizer that is assumed to be a self-concordant barrier; in our analysis, it is the loss functions that are assumed to be self-concordant barriers. The difference is illustrated by our improved analysis of EG. The entropy regularizer is not a self-concordant barrier, so the analyses of the interior-point methods, BSM, and LB-FTRL do not directly apply. However, EG naturally fits in our framework.

Indeed, BSM coincides with LB-FTRL with linearized losses. The original analysis of BSM by Nesterov [6] appears to be complicated to us. We provide an arguably simpler regret analysis of BSM, based on the approaches of Abernethy et al. [32] and van Erven et al. [26], in Appendix F.

# 3 Problem Formulations and Preliminaries

## 3.1 Online Convex Optimization

An online convex optimization problem is a multi-round game between two players, say Learner and Reality. Let  $\mathcal{X}$  be a convex set. In the  $t$ -th round, Learner announces some  $x_t \in \mathcal{X}$ , given the history  $\{x_1, \dots, x_{t-1}, f_1, \dots, f_{t-1}\}$ ; then, Reality announces a proper closed convex function  $f_t : \mathcal{X} \rightarrow (-\infty, +\infty]$ , given the history  $\{x_1, \dots, x_t, f_1, \dots, f_{t-1}\}$ . The loss suffered by Learner in the  $t$ -th round is given by  $f_t(x_t)$ . Let  $T \in \mathbb{N}$  be the time horizon. The *regret* is defined as

$$R_T(x) := \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x), \quad \forall x \in \mathcal{X}.$$

We will write  $R_T$ , without an input  $x$ , for the quantity  $\sup_{x \in \mathcal{X}} R_T(x)$ .

## 3.2 Online Portfolio Selection

Let  $d \in \mathbb{N}$ . OPS corresponds to an online convex optimization problem where

- the set  $\mathcal{X}$  is the probability simplex  $\Delta := \{(x(1), \dots, x(d)) \in \mathbb{R}_+^d \mid \sum_{i=1}^d x(i) = 1\}$  and
- the loss function  $f_t$  are given by  $f_t(x) := -\log \langle a_t, x \rangle$  for some  $a_t \in \mathbb{R}_+^d$ .

We assume that  $\|a_t\|_\infty \leq 1$ . If this is not the case, we can consider the sequence  $(\tilde{a}_t)_{t \in \mathbb{N}}$  for  $\tilde{a}_t := a_t / \|a_t\|_\infty$  instead of  $(a_t)_{t \in \mathbb{N}}$ ; the regret value is not affected.

**Interpretation** OPS characterizes a game of multi-round investment. Suppose there are  $d$  assets and Learner has  $w_1$  dollars. In the first round, Learner distributes the  $w_1$  dollars to the assets following the fractions given by  $x_1$ . The vector  $a_1$  denotes the price relatives of the  $d$  assets in that round. Then, after the first round, Learner has  $w_2 = w_1 \langle a_1, x_1 \rangle$  dollars. It is easily seen that after  $T$  rounds, we have

$$-\log \frac{w_{T+1}}{w_1} = \sum_{t=1}^{T+1} f_t(x_t).$$

A small cumulative loss implies a large wealth growth rate.

### 3.3 Quantum Generalization of OPS

The following three concepts are needed to understand the quantum generalization

1. A quantum state is represented by a *density matrix*, a Hermitian positive semi-definite matrix of unit trace. If there are  $q$  qubits, then the dimension of the density matrix is  $2^q \times 2^q$ . The vector of eigenvalues of a density matrix is a vector in the probability simplex.
2. A measurement is represented by a *positive operator-valued measure (POVM)*, a set of Hermitian positive semi-definite matrices summing up to the identity matrix. If we consider 1-by-1 matrices, i.e., real numbers, then a POVM also corresponds to a vector in the probability simplex.
3. Let  $\rho \in \mathbb{C}^{d \times d}$  be a density matrix and  $\{M_1, \dots, M_K\} \subset \mathbb{C}^{d \times d}$  be a POVM. The *measurement outcome* is a random variable  $\eta$  taking values in  $[K]$  such that

$$P(\eta = k) = \text{tr}(M_k \rho), \quad \forall k \in [K].$$

For convenience, we denote the set of density matrices in  $\mathbb{C}^{d \times d}$  by  $\mathcal{D}_d$ . We will denote density matrices by  $\rho$  and  $\sigma$  instead of  $x$  and  $y$ , following the convention in quantum information.

Let  $d \in \mathbb{N}$ . The quantum generalization of OPS corresponds to an online convex optimization problem where

- the set  $\mathcal{X}$  equals  $\mathcal{D}_d$  and
- the loss function  $f_t$  are given by  $f_t(\rho) := -\log \text{tr}(A_t \rho)$  for some  $A_t \in \mathbb{S}_+^d$ .

The quantum setup is challenging due to the presence of non-commutativity. When all matrices involved share the same eigenbasis, one may ignore the eigenbasis and focus on the vectors of eigenvalues. Then, it is easily seen that the generalization becomes equivalent to OPS. The reader is referred to, e.g., [15] and [16], for how the non-commutativity issue complicates the multiplicative weight update in the quantum case.

**Interpretation** Fix a POVM  $\{M_1, \dots, M_K\} \subset \mathbb{C}^{d \times d}$ . The class of all probability distributions on  $[K]$  associated with the POVM, parameterized by  $\mathcal{D}_d$ , is given by

$$\mathcal{P} := \{(\text{tr}(M_k \rho))_{k \in [K]} \mid \rho \in \mathcal{D}_d\}.$$

Consider the following problem of probability forecasting with the logarithmic loss [25, Chapter 9]. In the  $t$ -th round, Learner announces a probability distribution  $p_t = (p_t(k))_{k \in [K]} \in \mathcal{P}$ ; then, Reality announces some  $\eta_t \in [K]$ ; the loss suffered by Learner is given by  $-\log p_t(\eta_t)$ . It is easily checked that the probability forecasting problem is equivalent to the quantum generalization of OPS. Another closely related interpretation, based on maximum-likelihood quantum state tomography, was considered by [10].

### 3.4 Online Mirror Descent

We focus on OMD in this paper. Similar results to those in this paper can be derived also for FTRL with linearized losses. Consider the general online convex optimization problem in Section 3.1. Let  $h$  be a Legendre function such that the closure of  $\text{dom } h$  contains  $\mathcal{X}$ . Then,  $h$  is differentiable on  $\text{int dom } h$ . Define the associated Bregman divergence as

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle, \quad \forall (x, y) \in \text{dom } h \times \text{int dom } h. \quad (1)$$

OMD with  $h$  iterates as the following.

- Let  $x_1 \in \mathcal{X} \cap \text{int dom } h$ .

- For each  $t \in \mathbb{N}$ , compute

$$x_{t+1} = \arg \min_{x \in \mathcal{X}} \eta \langle \nabla f_t(x_t), x - x_t \rangle + D_h(x, x_t) \quad (2)$$

for some learning rate  $\eta > 0$ .

For OMD to be well defined, we always assume that  $\nabla f_t(x_t)$  exists for all  $t \in \mathbb{N}$ .

Below are two famous instances of OMD.

- OMD becomes online gradient descent when  $h = (1/2)\|\cdot\|_2^2$ .
- OMD becomes exponentiated gradient when  $h$  is the negative Shannon entropy and  $\mathcal{X}$  is the probability simplex.

### 3.5 Function Class

We will consider the class of relatively smooth self-concordant barrier loss functions.

**Definition 1** (Bauschke et al. [12], Lu et al. [13]). *Let  $\mathcal{X}$  be a convex set in  $\mathbb{R}^d$ . We say that a convex function  $f$  is  $L$ -smooth relative to a differentiable convex function  $h$  on  $\mathcal{X}$  for some  $L > 0$  if the function  $Lh - f$  is convex on  $\mathcal{X}$ .*

Relative smoothness becomes smoothness when  $h = (1/2)\|\cdot\|_2^2$ . Since the gradient of a convex function is a monotone mapping, the following lemma immediately follows.

**Lemma 2.** *Let  $f$  be a convex function  $L$ -smooth relative to a differentiable convex function  $h$  on  $\mathcal{X}$  for some  $L > 0$ . Then,*

$$L \langle \nabla h(y) - \nabla h(x), y - x \rangle \geq \langle \nabla f(y) - \nabla f(x), y - x \rangle, \quad \forall x, y \in \mathcal{X}.$$

**Definition 3** (Nesterov [33]). *We say a function  $f$  is  $M_f$ -self-concordant for some  $M_f \geq 0$  if*

$$|D^3 f(x)[u, u, u]| \leq 2M_f \langle u, \nabla^2 f(x)u \rangle^{3/2}, \quad \forall x \in \text{dom } f, u \in \mathbb{R}^d,$$

where

$$D^3 f(x)[u, u, u] := \left. \frac{d^3 f}{dt^3}(x + tu) \right|_{t=0}.$$

For a self-concordant function, the mononicity of its gradient mapping can be strengthened [33, Theorem 5.1.8].

**Lemma 4.** *Let  $f$  be an  $M_f$ -self-concordant function for some  $M_f \geq 0$ . Then,*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \frac{\|y - x\|_x^2}{1 + M_f \|y - x\|_x}, \quad \forall x, y \in \text{dom } f,$$

where  $\|y - x\|_x := \langle y - x, \nabla^2 f(x)(y - x) \rangle^{1/2}$ .

**Definition 5** (Nesterov [33]). *We say a convex function  $f$  is a  $\nu$ -self-concordant barrier if it is 1-self-concordant and*

$$\langle \nabla f(x), u \rangle^2 \leq \nu \langle u, \nabla^2 f(x)u \rangle, \quad \forall x \in \text{dom } f, u \in \mathbb{R}^d.$$

We now verify that the loss functions in OPS and its quantum generalization are relatively smooth self-concordant barriers. The following proposition is already proved by, e.g., Bauschke et al. [12, Lemma 7] and Nesterov [33, Example 5.3.1].

**Proposition 6.** *Consider the function  $f(x) = -\log \langle a, x \rangle$  for some  $a \in \mathbb{R}_+^d$ ,  $a \neq 0$ .*

1. *The function is 1-smooth relative to the logarithmic barrier on  $\mathbb{R}_{++}^d$ .*
2. *The function is a 1-self-concordant barrier.*

The proof of the following proposition is deferred to Appendix A.

**Proposition 7.** *Consider the function  $f(\rho) := -\log \text{tr}(A\rho)$  for some  $A \in \mathbb{S}_+^d$ ,  $A \neq 0$ .*

1. *The function is 1-smooth relative to the log-det function on  $\text{ri } \mathcal{D}_d$ .*
2. *The function is a 1-self-concordant barrier.*

## 4 Main Results

The results mentioned in Section 1 are presented in this section. Notice that there is a slight abuse of notations. The loss functions are always denoted as  $f_t$ , though they differ in different subsections; for OPS, the iterates are always denoted as  $x_t$ , though the iterates are generated by different algorithms in different subsections. For the quantum generalization of OPS, the iterates are denoted as  $\rho_t$  instead of  $x_t$ , following the convention of the quantum information community.

### 4.1 Online Self-Concordant and Relatively Smooth Minimization

Our study of OPS and its quantum generalization is based on the following general result.

**Theorem 8.** *Consider an online convex optimization problem where the loss functions are 1-self-concordant barriers and  $L$ -smooth relative to a Legendre function  $h$ . Then, OMD with  $h$  and learning rate  $\eta \in (0, 1/L)$  achieves*

$$R_T(x) \leq \frac{D_h(x, x_1)}{\eta} + \frac{TL\eta}{1 - L\eta}.$$

The proof of Theorem 8 is short, so we present it here. We start with the following lemma. The lemma should be familiar to convex optimization experts. Indeed, the lemma is obtained by simply adding time indices to the objective function in, e.g., the analysis by Lu et al. [13], Teboulle [34] for mirror descent minimizing convex relatively smooth functions. We provide a proof in Appendix B for completeness.

**Lemma 9.** *It holds that*

$$\sum_{t=1}^T f_t(x_{t+1}) - \sum_{t=1}^T f_t(x_t) \leq \frac{D_h(x, x_1)}{\eta}.$$

Notice that  $f_t(x_{t+1})$  is not realizable because computing  $x_{t+1}$  needs  $\nabla f_t$ . We then bound the difference between the actual regret and the “one-step-look-ahead” regret in Lemma 9.

**Lemma 10.** *It holds that*

$$\sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x_{t+1}) \leq \frac{TL\eta}{1 - L\eta}.$$

*Proof.* Define the local norm  $\|u\|_t := \langle u, \nabla^2 f_t(x_t) u \rangle^{1/2}$  and  $r_t := \|x_t - x_{t+1}\|_t$ . By the optimality condition of (2), we have

$$\langle \eta \nabla f_t(x_t) + \nabla h(x_{t+1}) - \nabla h(x_t), x_t - x_{t+1} \rangle \geq 0. \quad (3)$$

Then, we write

$$\begin{aligned} r_t &\geq \langle \nabla f_t(x_t), x_t - x_{t+1} \rangle \\ &\geq \frac{1}{\eta} \langle \nabla h(x_{t+1}) - \nabla h(x_t), x_{t+1} - x_t \rangle \\ &\geq \frac{1}{L\eta} \langle \nabla f_t(x_{t+1}) - \nabla f_t(x_t), x_{t+1} - x_t \rangle \\ &\geq \frac{r_t^2}{L\eta(1 + r_t)}. \end{aligned}$$

In the above, the first inequality follows from the definition of a self-concordant barrier (Definition 5); the second follows by rearranging (3); the third follows from the relative smoothness of  $f$ ; the fourth follows from Lemma 4. Solving the inequality for  $r_t$ , we get

$$r_t \leq \frac{L\eta}{1 - L\eta}.$$

Then, we write

$$f_t(x_t) - f_t(x_{t+1}) \leq \langle \nabla f_t(x_t), x_t - x_{t+1} \rangle \leq r_t \leq \frac{L\eta}{1 - L\eta}.$$

The lemma follows by summing the inequality from  $t = 1$  to  $t = T$ .  $\square$

Combining Lemma 9 and Lemma 10, Theorem 8 follows.

## 4.2 $\widetilde{\text{EG}}$

The original exponentiated gradient update requires the loss functions to be Lipschitz to achieve a sublinear regret. The Lipschitz assumption, as discussed in Section 1, does not hold in OPS and hence also its quantum generalization. Helmbold et al. [1] proposed a variant of the exponentiated gradient update, named  $\widetilde{\text{EG}}$ , that gets rid of the Lipschitz assumption in OPS. The  $\widetilde{\text{EG}}$  algorithm is presented in Algorithm 1, where  $e := (1, \dots, 1) \in \mathbb{R}^d$  and  $D_h$  denotes the Bregman divergence defined by  $h$  (1). Notice that  $\hat{x}_{t+1}$  admits the closed-form expression

$$\hat{x}_{t+1}(i) \propto \hat{x}_t(i) e^{-\eta \nabla_i \hat{f}_t(\hat{x}_t)}, \quad \forall i \in [d],$$

where  $\nabla_i \hat{f}_t(\hat{x}_t)$  denotes the  $i$ -th entry of  $\nabla \hat{f}_t(\hat{x}_t)$ .

---

**Algorithm 1**  $\widetilde{\text{EG}}$  for OPS.

---

**Input:**  $\eta > 0, \gamma \in (0, 1)$ .

- 1:  $x_1 \leftarrow e/d$ .
  - 2:  $\hat{x}_1 \leftarrow e/d$ .
  - 3:  $h(x) := \sum_{i=1}^d x(i) \log x(i) - \sum_{i=1}^d x(i)$ .
  - 4: **for all**  $t \in \mathbb{N}$  **do**
  - 5:    $\hat{a}_t \leftarrow (1 - \gamma/d)a_t + (\gamma/d)e$ .
  - 6:    $\hat{f}_t(x) := -\log \langle \hat{a}_t, x \rangle$ .
  - 7:    $\hat{x}_{t+1} \leftarrow \arg \min_{x \in \Delta} \eta \langle \nabla \hat{f}_t(\hat{x}_t), x - \hat{x}_t \rangle + D_h(x, \hat{x}_t)$ .
  - 8:    $x_{t+1} \leftarrow (1 - \gamma)\hat{x}_{t+1} + (\gamma/d)e$ .
  - 9: **end for**
- 

**Theorem 11.** *Suppose that  $T > 4d/\log d$ . Then, setting*

$$\gamma = \frac{2^{2/3}d^{1/3}}{(T \log d)^{1/3}} \quad \text{and} \quad \eta = \frac{\gamma\sqrt{d}}{\sqrt{Td\gamma + d\sqrt{\log d}}},$$

$\widetilde{\text{EG}}$  for OPS satisfies

$$\begin{aligned} R_T &\leq 2^{5/3}T^{2/3}d^{1/3}(\log d)^{2/3} + 2^{-2/3}T^{1/3}d^{2/3}(\log d)^{4/3} \\ &= \tilde{O}(T^{2/3}d^{1/3}). \end{aligned}$$

As stated in Section 1, the original regret bound for  $\widetilde{\text{EG}}$  is  $\tilde{O}(d^{1/2}T^{3/4})$ . The key to the improved regret bound is the following proposition.

**Proposition 12.** *For any  $t \in \mathbb{N}$ , the loss function  $f_t = -\log \langle a_t, x \rangle$  in OPS is  $G_t$ -smooth relative to the negative Shannon entropy  $h$  (see Algorithm 1) on  $\text{ri } \Delta$  for*

$$G_t := \sup_{x \in \Delta} \|\nabla f_t(x)\|_\infty = \max_{i,j \in [d]} \frac{a_t(i)}{a_t(j)},$$

where  $a_t(i)$  and  $a_t(j)$  denote the  $i$ -th and  $j$ -th entry of  $a_t$ , respectively.

A proof of the proposition above is provided in Appendix C. Though the proposition above seems to require the loss functions to be Lipschitz, the issue of non-Lipschitz losses in OPS has been handled by the specific form of  $\widetilde{\text{EG}}$ . Notice that  $\hat{G}_t := \sup_{x \in \Delta} \|\nabla \hat{f}_t(x)\|_\infty$  is upper bounded by  $d/\gamma$ . In view of Proposition 12, Theorem 8 can be applied to the sequences of points  $\{\hat{x}_t\}$  and loss functions  $\{\hat{f}_t\}$ , with relative smooth parameter  $d/\gamma$ . This explains why  $G_t$  does not appear in Theorem 11. The rest of the analysis is to estimate the difference between the regrets of  $\{x_t\}$  and  $\{\hat{x}_t\}$ . A proof of Theorem 11 is provided in Appendix D, which essentially follows Helmbold et al. [1].



---

**Algorithm 2** LB-OMD, online mirror descent with the logarithmic barrier, for OPS.

---

**Input:**  $\eta > 0$ .

- 1:  $h(x) := -\sum_{i=1}^d \log x(i)$ .
- 2:  $x_1 = e/d$ .
- 3: **for all**  $t \in \mathbb{N}$  **do**
- 4:    $x_{t+1} \leftarrow \arg \min_{x \in \Delta} \eta \langle \nabla f_t(x_t), x - x_t \rangle + D_h(x, x_t)$ .
- 5: **end for**

---

### 4.3 LB-OMD

LB-OMD is presented in Algorithm 2, where  $D_h$  denotes the Bregman divergence defined by  $h$ . The iterate  $x_{t+1}$  does not have a closed-form expression. Nevertheless, as pointed by Kotłowski and Neu [21, Appendix B],  $x_{t+1}$  can be efficiently computed by Newton’s method minimizing a self-concordant function on  $\mathbb{R}$ . The per-iteration time is hence  $\tilde{O}(d)$ .

**Theorem 13.** *Suppose that  $T > d$ . Setting*

$$\eta = \frac{\sqrt{d \log T}}{\sqrt{T} + \sqrt{d \log T}},$$

LB-OMD satisfies

$$R_T \leq 2\sqrt{Td \log T} + d \log T + 2 = \tilde{O}(\sqrt{Td}).$$

We will directly analyze the regret of a quantum generalization of LB-OMD (see the next subsection) in Appendix E. The proof Theorem 13 is similar and simpler and hence skipped.

### 4.4 Q-LB-OMD

Q-LB-OMD is presented in Algorithm 3, where  $I$  denotes the identity matrix and  $D_h$  denotes the Bregman divergence defined by  $h$ . Following the convention of quantum information, we denote the iterates by  $\rho_t$  instead of  $x_t$ . All matrices in the algorithm are of dimension  $d \times d$ . Q-LB-OMD is a direct quantum generalization of LB-OMD, in the sense that when all matrices involved share the same eigenbasis, then Q-LB-OMD is equivalent to LB-OMD. Similar to LB-OMD,  $\rho_{t+1}$  does not have a closed-form expression. Nevertheless, as pointed by Kotłowski and Neu [21, Appendix B], the iterate  $\rho_{t+1}$  can be computed by one eigendecomposition of  $\rho_t$  followed by Newton’s method minimizing a self-concordant function on  $\mathbb{R}$ . The per-iteration time is hence  $O(d^3)$ . In Section 1, we have shown that the per-iteration time of Q-LB-OMD is the shortest, in comparison to Q-Soft-Bayes and Schrödinger’s-BISONS.

---

**Algorithm 3** Q-LB-OMD, online mirror descent with the log-det function, for the quantum generalization of OPS.

---

**Input:**  $\eta > 0$ .

- 1:  $h(\rho) := -\log \det \rho$ .
- 2:  $\rho_1 = I/d$ .
- 3: **for all**  $t \in \mathbb{N}$  **do**
- 4:    $\rho_{t+1} \leftarrow \arg \min_{\rho \in \mathcal{D}_d} \eta \langle \nabla f_t(\rho_t), \rho - \rho_t \rangle + D_h(\rho, \rho_t)$ .
- 5: **end for**

---

**Theorem 14.** *The statement of Theorem 13 also holds for Q-LB-OMD.*

A proof of the theorem above is provided in Appendix E.

## 5 Concluding Remarks

We have achieved the following.

1. Provide an improved regret bound of  $\widetilde{\text{EG}}$  for OPS.
2. Show that LB-OMD is on the current efficiency-regret Pareto frontier for OPS.

3. Show that Q-LB-OMD inherits the regret bound of LB-OMD and hence, regarding its scalability with respect to the dimension in both regret and per-iteration time, is competitive among existing algorithms.

The key idea in our analyses is to exploit the self-concordant barrier property and relative smoothness of the loss functions simultaneously. We are not aware of any literature that adopts this approach. In this paper, we demonstrate the benefit of this approach with a standard algorithm: OMD with a constant learning rate. This approach may help us develop and analyze other, perhaps more complicated, algorithms for OPS and its quantum generalization.

The regret bounds in this paper are likely to be sub-optimal for the problem class we consider: online convex optimization with losses that are relatively smooth and self-concordant barriers. Indeed, it is easily checked that a self-concordant barrier is necessarily exp-concave [33]. Therefore, for the problem class we consider, logarithmic regrets may be achieved by the exponentially weighted online optimization (EWO) algorithm [35]. For example, EWO coincides with UPS, known to be regret-optimal, for online portfolio selection. Zimmert et al. [9] claimed that EWO achieves a  $O(d^2 \log T)$  regret for online learning quantum states with the logarithmic loss, but we have not found a proof.

Recall that, as discussed in the introduction, our aim is not to seek for regret-optimal algorithms, but to strike a balance between efficiency and regret. EWO requires evaluating the expectation of a data-determined probability distribution in each iteration, which is computationally very expensive. The per-iteration time of UPS is already formidable. We are not aware of any polynomial-time implementation of EWO for online learning quantum states with the logarithmic loss.

## Acknowledgments

We thank the anonymous reviewers for their inspiring comments.

C.-E. Tsai and Y.-H. Li are supported by the Young Scholar Fellowship (Einstein Program) of the National Science and Technology Council of Taiwan under grant numbers MOST 108-2636-E-002-014, MOST 109-2636-E-002-025, MOST 110-2636-E-002-012, MOST 111-2636-E-002-019, and NSTC 112-2636-E-002-003 and by the research project ‘‘Pioneering Research in Forefront Quantum Computing, Learning and Engineering’’ of National Taiwan University under grant numbers NTU-CC-111L894606 and NTU-CC-112L893406.

H.-C. Cheng is supported by the Young Scholar Fellowship (Einstein Program) of the National Science and Technology Council (NSTC) in Taiwan (R.O.C.) under Grant MOST 111-2636-E-002-026, Grant MOST 111-2119-M-007-006, Grant MOST 111-2119-M-001-004, and is supported by the Yushan Young Scholar Program of the Ministry of Education in Taiwan (R.O.C.) under Grant NTU-111V1904-3, and Grant NTU-111L3401, and by the research project ‘‘Pioneering Research in Forefront Quantum Computing, Learning and Engineering’’ of National Taiwan University under Grant No. NTU-CC-111L894605.’’

## References

- [1] David P. Helmbold, Robert E. Shapire, Yoram Singer, and Manfred K. Warmuth. On-line portfolio selection using multiplicative updates. *Math. Financ.*, 8(4):325–347, 1998.
- [2] Laurent Orseau, Tor Lattimore, and Shane Legg. Soft-Bayes: Prod for mixtures of experts with log-loss. In *Proc. 28th Int. Conf. Algorithmic Learning Theory*, pages 372–399, 2017.
- [3] Thomas M. Cover. Universal portfolios. *Math. Financ.*, 1(1):1–29, 1991.
- [4] Thomas M. Cover and Erick Ordentlich. Universal portfolios with side information. *IEEE Trans. Inf. Theory*, 42(2):348–363, 1996.
- [5] Adam Kalai and Santosh Vempala. Efficient algorithms for universal portfolios. *J. Mach. Learn. Res.*, 3:423–440, 2002.
- [6] Yurii Nesterov. Barrier subgradient method. *Math. Program., Ser. B*, 127:31–56, 2011.

- [7] Haipeng Luo, Chen-Yu Wei, and Kai Zheng. Efficient online portfolio with logarithmic regret. In *Adv. Neural Information Processing Systems 31*, 2018.
- [8] Zakaria Mhammedi and Alexander Rakhlin. Damped Online Newton Step for Portfolio Selection. In *Proc. 35th Annu. Conf. Learning Theory*, pages 5561–5595, 2022.
- [9] Julian Zimmert, Naman Agarwal, and Satyen Kale. Pushing the efficiency-regret Pareto frontier for online learning of portfolios and quantum states. In *Proc. 35th Annu. Conf. Learning Theory*, pages 182–226, 2022.
- [10] Chien-Ming Lin, Yu-Ming Hsu, and Yen-Huan Li. An Online Algorithm for Maximum-Likelihood Quantum State Tomography. In *24th Annu. Conf. Quantum Information Processing*, 2021. arXiv:2012.15498.
- [11] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, PA, 1994.
- [12] Heinz H. Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Math. Oper. Res.*, 42(2): 330–348, 2017.
- [13] Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM J. Optim.*, 28(1):333–354, 2018.
- [14] Nima Eshraghi and Ben Liang. Dynamic regret of online mirror descent for relatively smooth convex cost functions. *IEEE Control Syst. Lett.*, 6:2395–2400, 2022.
- [15] Manfred K. Warmuth and Dima Kuzmin. Online variance minimization. In *19th Annu. Conf. Learning Theory, COLT 2006*, pages 514–528, 2006.
- [16] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *STOC '07: Proc. 39th Annu. ACM Symp. Theory of computing*, pages 227–236, 2007.
- [17] Wouter M. Koolen, Wojtek Kotłowski, and Manfred K. Warmuth. Learning eigenvectors for free. In *Adv. Neural Information Processing Systems 24*, 2011.
- [18] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. In *Adv. Neural Information Processing Systems 31*, 2018.
- [19] Xinyi Chen, Elad Hazan, Tongyang Li, Zhou Lu, Xinzhao Wang, and Rui Yang. Adaptive online learning of quantum states. 2022. arXiv:2206.00220 [cs.LG].
- [20] Feidiao Yang, Jiaqing Jiang, Jialin Zhang, and Xiaoming Sun. Revisiting online quantum state learning. In *Proc. AAAI Conf. Artificial Intelligence*, 2020.
- [21] Wojciech Kotłowski and Gergely Neu. Bandit principal component analysis. In *Proc. 32nd Conf. Learning Theory*, pages 1994–2024, 2019.
- [22] Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proc. 21st Annu. Conf. Learning Theory*, pages 263–273, 2008.
- [23] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proc. 26th Annu. Conf. Learning Theory*, pages 993–1019, 2013.
- [24] Elad Hazan. Introduction to online convex optimization. *Found. Trends Opt.*, 2(3–4):157–325, 2016.
- [25] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge Univ. Press, Cambridge, UK, 2006.
- [26] Tim van Erven, Dirk van der Hoeven, Wojciech Kotłowski, and Wouter M. Koolen. Open Problem: Fast and Optimal Online Portfolio Selection. In *Proc. 33rd Annu. Conf. Learning Theory*, pages 3864–3869, 2020.

- [27] Lijun Zhang, Tianbao Yang, Jinfeng Yi, Rong Jin, and Zhi-Hua Zhou. Improved dynamic regret for non-degenerate functions. In *Adv. in Neural Information Processing Systems 30*, 2017.
- [28] Nathan Srebro, Karthik Sridharan, and Ambuj Tewari. Smoothness, low-noise and fast rates. In *Adv. Neural Information Processing Systems 23*, 2010.
- [29] Francesco Orabona. A modern introduction to online learning. 2022. arXiv:1912.13213v5 [cs.LG].
- [30] Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Proc. 18th Int. Conf. Artificial Intelligence and Statistics*, pages 398–406, 2015.
- [31] Aryan Mokhtari, Shahin Shahrampour, Ali Jadbabaie, and Alejandro Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *IEEE 55th Conf. Decision and Control (CDC)*, pages 7195–7201, 2016.
- [32] Jacob D. Abernethy, Elad Hazan, and Alexander Rakhlin. Interior-Point Methods for Full-Information and Bandit Online Learning. *IEEE Trans. Inf. Theory*, 58(7):4164–4175, 2012.
- [33] Yurii Nesterov. *Lectures on Convex Optimization*. Springer, Cham, CH, second edition, 2018.
- [34] Marc Teboulle. A simplified view of first order methods for optimization. *Math. Program., Ser. B*, 170:67–96, 2018.
- [35] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Mach. Learn.*, 69(2):169–192, 2007.
- [36] Rajendra Bhatia. *Matrix Analysis*. Springer, New York, NY, 1997.
- [37] Fumio Hiai and Dénes Petz. *Introduction to Matrix Analysis and Applications*. Springer, Cham, 2014.

## A Proof of Proposition 7

### A.1 Relative Smoothness

We will use the following lemma [36, Exercise IV.2.7].

**Lemma 15** (A matrix Cauchy-Schwarz inequality). *Let  $K, L \in \mathbb{C}^{d \times d}$ . Then,*

$$(\operatorname{tr} |KL|)^2 \leq \operatorname{tr}(|K|^2) \operatorname{tr}(|L|^2)$$

where  $|A| := (A^*A)^{1/2}$ .

Define  $\varphi(\rho) := \log \operatorname{tr}(A\rho) - \log \det \rho$ . It suffices to show that  $\varphi$  is convex on  $\operatorname{ri} \mathcal{D}_d$ ; equivalently, it suffices to show that  $D^2\varphi(\rho)[\sigma, \sigma] \geq 0$  for all  $\rho \in \operatorname{ri} \mathcal{D}_d$  and  $\sigma \in \mathbb{S}^d$  [12, Proposition 17.7]. Since  $\varphi(\rho) = \log \operatorname{tr}(A\rho) - \operatorname{tr} \log(\rho)$  [37, Theorem 3.13], we write [37, Theorem 3.23 and Example 3.20]

$$\begin{aligned} D^2\varphi(\rho)[\sigma, \sigma] &= \left. \frac{d^2\varphi}{dt^2}(\rho + t\sigma) \right|_{t=0} \\ &= \left. \frac{d}{dt} \left\{ \frac{\operatorname{tr}(A\sigma)}{\operatorname{tr}(A\rho + tA\sigma)} - \operatorname{tr}[\sigma(\rho + t\sigma)^{-1}] \right\} \right|_{t=0} \\ &= \left. \left\{ - \left( \frac{\operatorname{tr}(A\sigma)}{\operatorname{tr}(A\rho + tA\sigma)} \right)^2 + \operatorname{tr}[\sigma(\rho + t\sigma)^{-1}\sigma(\rho + t\sigma)^{-1}] \right\} \right|_{t=0} \\ &= - \left( \frac{\operatorname{tr}(A\sigma)}{\operatorname{tr}(A\rho)} \right)^2 + \operatorname{tr}[(\sigma\rho^{-1})^2]. \end{aligned}$$

Now, it suffices to prove that  $(\operatorname{tr}(A\sigma))^2 \leq (\operatorname{tr}(A\rho))^2 \operatorname{tr}((\sigma\rho^{-1})^2)$ . Let  $K = \rho^{1/2}A\rho^{1/2}$  and  $L = \rho^{-1/2}\sigma\rho^{-1/2}$ . Applying Lemma 15 with  $K = \rho^{1/2}A\rho^{1/2}$  and  $L = \rho^{-1/2}\sigma\rho^{-1/2}$ , we have

$$\begin{aligned} (\operatorname{tr}(A\sigma))^2 &= (\operatorname{tr}(KL))^2 \\ &\leq (\operatorname{tr} |KL|)^2 \\ &\leq \operatorname{tr}(|K|^2) \operatorname{tr}(|L|^2) \\ &= \operatorname{tr}((A\rho)^2) \operatorname{tr}((\sigma\rho^{-1})^2) \\ &\leq (\operatorname{tr}(A\rho))^2 \operatorname{tr}((\sigma\rho^{-1})^2). \end{aligned}$$

In the above, the first inequality follows from Weyl's majorant theorem [36, Theorem II.3.6]; the second inequality follows from Lemma 15. As for the third inequality, because  $A, \rho \in \mathbb{S}_+^D$ ,  $A\rho$  has non-negative eigenvalues  $\lambda_i \geq 0$  ( $i \in [D]$ ). Then,

$$\operatorname{tr}((A\rho)^2) = \sum_{i=1}^d \lambda_i^2 \leq \left( \sum_{i=1}^d \lambda_i \right)^2 = (\operatorname{tr}(A\rho))^2.$$

This completes the proof.

### A.2 Self-Concordant Barrier Property

First,  $\operatorname{dom} f = \{\rho \in \mathbb{S}^d \mid \operatorname{tr}(A\rho) > 0\}$  is open in  $\mathbb{S}^d$ . A direct calculation gives

$$\begin{aligned} Df(\rho)[\sigma] &= -\frac{\operatorname{tr}(A\sigma)}{\operatorname{tr}(A\rho)}, \quad D^2f(\rho)[\sigma, \sigma] = \left( \frac{\operatorname{tr}(A\sigma)}{\operatorname{tr}(A\rho)} \right)^2 \\ D^3f(\rho)[\sigma, \sigma, \sigma] &= -2 \left( \frac{\operatorname{tr}(A\sigma)}{\operatorname{tr}(A\rho)} \right)^2. \end{aligned}$$

The self-concordant barrier property of  $f$  follows.

## B Proof of Lemma 9

We will use the following lemma [34, Lemma 3.1].

**Lemma 16** (Bregman proximal inequality). *Let  $\varphi : \mathbb{R}^d \rightarrow (-\infty, \infty]$  be a proper closed convex function. Let  $x \in \text{int dom } h$  and  $\eta > 0$ . Define*

$$x_+ \in \arg \min_{u \in \mathbb{R}^d} \varphi(u) + \frac{D_h(u, x)}{\eta}.$$

Then,

$$\varphi(x_+) - \varphi(u) \leq \frac{1}{\eta} (D_h(u, x) - D_h(u, x_+) - D_h(x_+, x)), \quad \forall u \in \text{dom } h.$$

By the  $L$ -smoothness of  $f_t$  relative to  $h$  and convexity of  $f_t$ , we write

$$\begin{aligned} f_t(x_{t+1}) &\leq f_t(x_t) + \langle \nabla f_t(x_t), x_{t+1} - x_t \rangle + LD_h(x_{t+1}, x_t) \\ &= f_t(x_t) + \langle \nabla f_t(x_t), x - x_t \rangle + \langle \nabla f_t(x_t), x_{t+1} - x \rangle + LD_h(x_{t+1}, x_t) \\ &\leq f_t(x) + \langle \nabla f_t(x_t), x_{t+1} - x \rangle + LD_h(x_{t+1}, x_t). \end{aligned}$$

Applying Lemma 16 to (2), we write

$$\langle \nabla f_t(x_t), x_{t+1} - x \rangle \leq \frac{1}{\eta} (D_h(x, x_t) - D_h(x, x_{t+1}) - D_h(x_{t+1}, x_t)).$$

Combining the two inequalities, we get

$$\begin{aligned} f_t(x_{t+1}) &\leq f_t(x) + \frac{D_h(x, x_t)}{\eta} - \frac{D_h(x, x_{t+1})}{\eta} + \left(L - \frac{1}{\eta}\right) D_h(x_{t+1}, x_t) \\ &\leq f_t(x) + \frac{D_h(x, x_t)}{\eta} - \frac{D_h(x, x_{t+1})}{\eta}, \end{aligned}$$

where the last line is by the assumption  $\eta < 1/L$ . The lemma then follows by a telescopic sum.

## C Proof of Proposition 12

We drop the subscript  $t$  in  $f_t$ ,  $a_t$ , and  $G_t$  for convenience. It suffices to show that  $\nabla^2 f(x) \leq G \nabla^2 h(x)$  for all  $x \in \text{ri } \Delta$ . We write

$$\begin{aligned} \langle v, \nabla^2 f(x)v \rangle &= \left( \frac{\langle a, v \rangle}{\langle a, x \rangle} \right)^2 \\ &= \left( \sum_{i=1}^d \frac{a(i)x(i)}{\langle a, x \rangle} \frac{v(i)}{x(i)} \right)^2 \\ &\leq \sum_{i=1}^d \frac{a(i)x(i)}{\langle a, x \rangle} \left( \frac{v(i)}{x(i)} \right)^2 \\ &= \sum_{i=1}^d \frac{a(i)}{\langle a, x \rangle} \frac{(v(i))^2}{x(i)} \\ &\leq G \sum_{i=1}^d \frac{(v(i))^2}{x(i)}, \end{aligned}$$

where the first inequality follows from Jensen's inequality and the second follows from the definition of  $G$ . It remains to notice that

$$\sum_{i=1}^d \frac{(v(i))^2}{x(i)} = \langle v, \nabla^2 h(x)v \rangle.$$

## D Proof of Theorem 11

The proof strategy is similar to that of Helmbold et al. [1, proof of Theorem 4.2]. Since  $\|a_t\|_\infty \leq 1$ ,  $\|\hat{a}_t\|_\infty \leq 1$ . Then, by the definition of  $\hat{a}_t$ , we have

$$\hat{G}_\infty := \sup_{t \in \mathbb{N}} \max_{i, j \in [d]} \frac{\hat{a}_t(i)}{\hat{a}_t(j)} \leq \frac{d}{\gamma}.$$

Notice that  $D_h(x, x_1) \leq \log d$  for all  $x \in \Delta$ . Theorem 8 then implies

$$\sum_{t=1}^T \hat{f}_t(\hat{x}_t) - \sum_{t=1}^T \hat{f}_t(x) \leq \frac{\log d}{\eta} + \frac{Td\eta}{\gamma - d\eta}. \quad (4)$$

By the definition of  $\hat{a}_t$ , we have

$$\langle \hat{a}_t, x \rangle = \left(1 - \frac{\gamma}{d}\right) \langle a_t, x \rangle + \frac{\gamma}{d} \geq \langle a_t, x \rangle.$$

Therefore,

$$\sum_{t=1}^T \hat{f}_t(x) \leq \sum_{t=1}^T f_t(x). \quad (5)$$

Following the argument before (4.4) of Helmbold et al. [1, proof of Theorem 4.2], we have

$$\log \langle a_t, x_t \rangle \geq \log \langle \hat{a}_t, \hat{x}_t \rangle + \log \left(1 - \gamma + \frac{\gamma}{d}\right).$$

By Jensen's inequality<sup>2</sup>,

$$\log \left(1 - \gamma + \frac{\gamma}{d}\right) \geq (1 - \gamma) \log 1 + \gamma \log \frac{1}{d} = -\gamma \log d.$$

Therefore,

$$\sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T \hat{f}_t(\hat{x}_t) \leq \gamma T \log d. \quad (6)$$

Combining (4), (5), (6), we get

$$R_T \leq \frac{\log d}{\eta} + \frac{Td\eta}{\gamma - d\eta} + \gamma T \log d.$$

We first choose

$$\eta = \frac{\gamma \sqrt{\log d}}{\sqrt{Td\gamma} + d\sqrt{\log d}},$$

which gives

$$R_T \leq \frac{2\sqrt{Td\log d}}{\sqrt{\gamma}} + \frac{d \log d}{\gamma} + \gamma T \log d.$$

We then choose  $\gamma$  that minimizes the sum of the first and third terms. The constraint on  $T$  is to ensure that  $\gamma \in (0, 1)$ .

## E Proof of Theorem 14

Since  $D_h(x, I/d)$  can be arbitrarily large when  $x$  approaches the boundary of  $\mathcal{D}_d$ , we cannot directly apply Theorem 8. We make use of the following lemma [7, Lemma 10].

**Lemma 17.** *Define*

$$\bar{\rho} := \left(1 - \frac{1}{T}\right) \rho + \frac{I}{Td}, \quad \forall \rho \in \mathcal{D}_d.$$

*Then,*

$$R_T(\rho) \leq R_T(\bar{\rho}) + 2.$$

<sup>2</sup>Helmbold et al. [1, proof of Theorem 4.2] assumes  $\gamma \in (0, 1/2]$  and bounds the quantity above by  $-2\gamma$ .

*Proof.* The proof of Luo et al. [7, Lemma 10] only requires the convexity of  $f_t$  and the fact that  $\text{tr}(A_t \rho) \leq \text{tr}(A_t \bar{\rho}/(1 - 1/T))$  (in our context) and hence directly extends for the quantum case.  $\square$

By Lemma 17, we have

$$\begin{aligned} R_T(\rho) &\leq R_T(\bar{\rho}) + 2 \\ &\leq \frac{D_h(\bar{\rho}, I/d)}{\eta} + \frac{T\eta}{1-\eta} + 2. \end{aligned}$$

It remains to notice that

$$D_h(\bar{\rho}, I/d) = -\log \det \bar{\rho} - d \log d = -\text{tr}(\log \bar{\rho}) - d \log d \leq d \log T.$$

## F Alternative Analysis of LB-FTRL with Linearized Losses

---

**Algorithm 4** LB-FTRL with linearized losses

---

**Input:**  $\eta > 0$ .  
1:  $h(x) := -\sum_{i=1}^d \log x(i)$ .  
2:  $x_1 \in \arg \min_{x \in \Delta} h(x)$ .  
3: **for all**  $t \in \mathbb{N}$  **do**  
4:  $x_{t+1} \leftarrow \arg \min_{x \in \Delta} \eta \sum_{\tau=1}^t \langle \nabla f_\tau(x_\tau), x - x_\tau \rangle + h(x)$ .  
5: **end for**

---



---

**Algorithm 5** FTRL with self-concordant barrier regularizer

---

**Input:**  $\eta > 0$ , a constraint set  $\mathcal{X}$ , and a self-concordant barrier  $R$  for  $\mathcal{X}$ .  
1:  $x_1 \in \arg \min_{x \in \mathcal{X}} R(x)$ .  
2: **for all**  $t \in \mathbb{N}$  **do**  
3:  $x_{t+1} \leftarrow \arg \min_{x \in \mathcal{X}} \eta \sum_{\tau=1}^t \langle \nabla f_\tau(x_\tau), x - x_\tau \rangle + R(x)$ .  
4: **end for**

---

This section provides an arguably simpler analysis of LB-FTRL with linearized losses (Algorithm 4) for OPS, showing that the regret of Algorithm 4 is  $\tilde{O}(\sqrt{dT})$ .

We start with the following regret bound for FTRL with a self-concordant barrier regularizer due to Abernethy et al. [32].

**Theorem 18** (Abernethy et al. [32]). *Let  $\|\cdot\|_x$  be the local norm associated with  $R$  at the point  $x$ . Assume that  $\eta \|\nabla f_t(x_t)\|_{x_t}^{*2} \leq 1/4$  for all  $t \in \mathbb{N}$ . Then, for any  $x \in \mathcal{X}$ , Algorithm 5 satisfies*

$$R_T(x) \leq \frac{R(x) - R(x_1)}{\eta} + 2\eta \sum_{t=1}^T \|\nabla f_t(x_t)\|_{x_t}^{*2}, \quad \forall x \in \mathcal{X}.$$

The logarithmic barrier  $h$  is not a self-concordant barrier for  $\Delta$ , so Theorem 18 does not directly apply to Algorithm 4. Nevertheless, there is a trick to handle the issue. As van Erven et al. [26] and Mhammedi and Rakhlin [8] do, consider an affine transformation  $\Pi : \mathcal{W} \rightarrow \Delta$  defined as follows.

$$\begin{aligned} \mathcal{W} &:= \left\{ w \in \mathbb{R}_+^{d-1} : w_i \geq 0 \ \forall i \in [d-1], \quad \sum_{i=1}^{d-1} w_i \leq 1 \right\}, \\ \Pi(w) &= Aw + u_d, \quad A = \begin{pmatrix} I_{d-1} \\ -e^\top \end{pmatrix}, \quad u_d = (0, \dots, 0, 1)^\top, \end{aligned}$$

where with a slight abuse of notation,  $e$  denotes the all-one vector in  $\mathbb{R}^{d-1}$ . The transformation



$\Pi$  is a bijection between  $\mathcal{W}$  and  $\Delta$ . Then, Algorithm 4 can be equivalently written as

$$\begin{aligned} w_1 &\in \arg \min_{w \in \mathcal{W}} \tilde{h}(w), \\ w_{t+1} &\in \arg \min_{w \in \mathcal{W}} \eta \sum_{\tau=1}^t \langle \nabla \tilde{f}_\tau(w_\tau), w - w_\tau \rangle + \tilde{h}(w), \quad \forall t \geq 1, \\ x_t &= \Pi(w_t), \quad \forall t \geq 1, \end{aligned}$$

where  $\tilde{f}_\tau = f_\tau \circ \Pi$  and  $\tilde{h} = h \circ \Pi$ . It is easily checked that  $\tilde{h}$  is a self-concordant barrier for  $\mathcal{W}$ . We then obtain the following result, as a corollary of Theorem 18.

**Theorem 19.** *Let  $\|\cdot\|_x$  be the local norm associated with  $h$  at the point  $x$ . Assume that  $\eta \|\nabla f_t(x_t)\|_{x_t}^{*2} \leq 1/4$  for all  $t \in \mathbb{N}$ . Then, Algorithm 4 satisfies*

$$R_T(x) \leq \frac{h(x) - h(x_1)}{\eta} + 2\eta \sum_{t=1}^T \|\nabla f_t(x_t)\|_{x_t}^{*2}, \quad \forall x \in \Delta.$$

*Proof.* By the equivalence mentioned above, Theorem 18 implies

$$\sum_{t=1}^T \tilde{f}_t(w_t) - \tilde{f}_t(w) \leq \frac{\tilde{h}(w) - \tilde{h}(w_1)}{\eta} + 2\eta \sum_{t=1}^T \left\| \nabla \tilde{f}_t(w_t) \right\|_{w_t}^{*2}, \quad \forall w \in \mathcal{W},$$

where  $\|\cdot\|_w$  is the local norm associated with  $\tilde{h}$  at  $w$ . It remains to notice that

$$\left\| \nabla \tilde{f}_t(w_t) \right\|_{w_t}^{*2} = \langle \nabla f_t(x_t), A(A^\top \nabla^2 h(x_t)A)^{-1} A^\top \nabla f_t(x_t) \rangle \leq \|\nabla f_t(x_t)\|_{x_t}^{*2}.$$

Then, the theorem follows.  $\square$

The following fact was used in [26]. We provide a proof for completeness.

**Lemma 20.** *It holds that  $\|\nabla f_t(x_t)\|_{x_t}^{*2} \leq 1$  for all  $t \in \mathbb{N}$ .*

*Proof.* By Proposition 6, we have  $\nabla f_t(x_t) \nabla f_t(x_t)^\top = \nabla^2 f_t(x_t) \leq \nabla^2 h(x_t)$ . For any  $\varepsilon > 0$ ,

$$\begin{aligned} &\langle \nabla f_t(x_t), (\varepsilon I + \nabla^2 h(x_t))^{-1} \nabla f_t(x_t) \rangle \\ &\leq \langle \nabla f_t(x_t), (\varepsilon I + \nabla f_t(x_t) \nabla f_t(x_t)^\top)^{-1} \nabla f_t(x_t) \rangle \\ &= \frac{\|\nabla f_t(x_t)\|_2^2}{\varepsilon + \|\nabla f_t(x_t)\|_2^2}. \end{aligned}$$

Letting  $\varepsilon \rightarrow 0$  completes the proof.  $\square$

By Theorem 19, Lemma 20, and Lemma 17, we obtain the following.

**Corollary 21.** *For  $\eta \leq 1/4$ , Algorithm 4 satisfies*

$$R_T(x) \leq \frac{d \log T}{\eta} + 2\eta T + 2, \quad \forall x \in \Delta.$$

Optimizing over  $\eta$ , we get the desired  $\tilde{O}(\sqrt{dT})$  regret bound.