Proceedings of Machine Learning Research vol 178:1–45, 2022 35th Annual Conference on Learning Theory

Pushing the Efficiency-Regret Pareto Frontier for Online Learning of Portfolios and Quantum States

Julian Zimmert ZIMMERT@GOOGLE.COM
Naman Agarwal NAMANAGARWAL@GOOGLE.COM
Satyen Kale SATYENKALE@GOOGLE.COM
Google Research

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We revisit the classical online portfolio selection problem. It is widely assumed that a trade-off between computational complexity and regret is unavoidable, with Cover's Universal Portfolios algorithm, SOFT-BAYES and ADA-BARRONS currently constituting its state-of-the-art Pareto frontier. In this paper, we present the first efficient algorithm, BISONS, that obtains polylogarithmic regret with memory and per-step running time requirements that are polynomial in the dimension, displacing ADA-BARRONS from the Pareto frontier. Additionally, we resolve a COLT 2020 open problem by showing that a certain Follow-The-Regularized-Leader algorithm with log-barrier regularization suffers an exponentially larger dependence on the dimension than previously conjectured. Thus, we rule out this algorithm as a candidate for the Pareto frontier. We also extend our algorithm and analysis to a more general problem than online portfolio selection, viz. online learning of quantum states with log loss. This algorithm, called SCHRÖDINGER'S-BISONS, is the first efficient algorithm with polylogarithmic regret for this more general problem.

Keywords: Portfolio Management, Online Learning, Quantum Learning

1. Introduction

We study the classical online portfolio selection problem (Cover, 1991). In this problem, there are d assets (e.g. stocks) that an investor can invest money in on any given day. On each day, indexed by $t = 1, 2, \dots, T$, the investor can choose a *portfolio* over the d assets, which is a distribution of their wealth on the assets, after observing the *returns* (i.e. ratio of closing price to opening price) of the assets on the previous day. The goal is to compete with the best *constant-rebalanced portfolio* (CRP) in hindsight, which redistributes wealth on each day to maintain a fixed proportion in each asset. Importantly, we study the case without assumptions on the quality of the returns, i.e. any individual asset might suffer a total loss at any time. On any day, the wealth of the investor increases by a factor equal to the inner product between the portfolio chosen by the investor and the vector of returns for the d assets. The goal is to develop algorithms that minimize the investor's *regret*, which is the difference between the *logarithm* of the total wealth earned by the investor after T days (starting with an initial wealth of \$1), and the logarithm of the total wealth earned by the best CRP in hindsight. Equivalently, the online portfolio selection problem can be seen as an instance of online convex optimization (OCO), where the loss is the negative logarithm of the inner product between the portfolio and the returns vector.

The online portfolio selection problem can be seen as a special case of a more general problem, viz. online learning of quantum states with log loss. In this problem, the goal is to learn to predict the outcome of a sequence of *two-outcome measurements* of an unknown quantum state on $\log_2(d)$

qubits. Without going into quantum computing jargon (we refer the reader to (Aaronson et al., 2018) and Appendix B.1 for a more detailed discussion of the setting), this online learning problem can be specified as follows. In each time step the learner constructs a quantum state, which is a $d \times d$ positive semidefinite Hermitian matrix of trace 1, and in response, receives a *two-outcome measurement*, which is a $d \times d$ Hermitian matrix with eigenvalues in $[0, 1]$ [1]. The loss of the learner is the negative logarithm of the trace product between the quantum state generated by the learner and the measurement. The trace product can be interpreted as a probabilistic prediction of observing one of two outcomes in the measurement, and hence it is natural to use the log loss for measuring the quality of the prediction. The goal is to minimize regret with respect to the best quantum state in hindsight. It is easy to see that the online portfolio selection problem is exactly the special case of this problem where both the quantum state and loss matrices are restricted to be diagonal matrices. Aaronson et al. (2018) developed regret minimizing algorithms for *Lipschitz* loss functions of the trace product – in particular, the natural log loss setting was not handled by their algorithms.

Our first main contribution is the development of new algorithms, BISONS for the online portfolios problem and SCHRÖDINGER'S-BISONS for the quantum learning problem, with regret bounds of $\mathcal{O}(d^2 \log^2(T))$ and $\mathcal{O}(d^3 \log^2(T))$ respectively, and $\tilde{\mathcal{O}}(\text{poly}(d))$ [2] per-iteration running time. This result is noteworthy for two reasons. BISONS is the first algorithm that enjoys polylogarithmic regret with $\tilde{\mathcal{O}}(\text{poly}(d))$ memory and running time per-iteration, and we show that the quantum learning problem is only slightly harder than the online portfolios problem. Technically, the BISONS algorithm operates in epochs (inspired by the ADA-BARRONS algorithm of Luo et al. (2018)), with each epoch running a Follow-The-Regularized Leader (FTRL) algorithm with quadratic surrogate losses using the log-barrier regularizer, with an additional *linear bias* term added to the surrogate loss. The linear bias term is crucial to the analysis and ensures that the regret within any epoch is non-positive, while the final epoch incurs polylogarithmic regret.

Extending the algorithm and its analysis to the quantum learning problem presents several technical challenges. First, the non-commutativity of the matrices involved makes the construction of the linear bias term non-trivial; we use semidefinite programming duality to design the linear term. Second, since the matrices are *complex* and Hermitian, standard convex analysis machinery such as gradients, Hessians and the intermediate value theorem need to be custom developed for the analysis. As observed earlier, the portfolios problem is a special case of the quantum learning problem when the matrices are all diagonal, and in this case SCHRÖDINGER'S-BISONS collapses to BISONS. Hence, we only give a regret bound analysis for SCHRÖDINGER'S-BISONS using the machinery developed; the bound for BISONS follows easily accounting for the dimension of the problem.

Our second main contribution is that we provide novel insights about a certain natural FTRL algorithm for the online portfolios problem. Van Erven et al. (2020) conjectured, in a COLT 2020 open problem, that FTRL with log-barrier regularization (denoted LB-FTRL) obtains the optimal $\mathcal{O}(d \log(T))$ regret bound. If this were true, this would provide the first (semi-)efficient algorithm with optimal regret. We resolve the COLT 2020 open problem by disproving this conjecture with a

1. It may not be apparent that the portfolio selection problem is a special case of the quantum learning problem since we didn't assume that returns in the online portfolio selection problem lie in $[0, 1]$. But since the regret in the portfolio problem is invariant to multiplicative scaling of the returns vectors, we may indeed assume this without loss of generality. We do caution the reader that in the description that follows, we will actually assume a different (but equivalent) normalization for technical reasons: i.e., the returns sum up to 1, and the trace of the measurement matrices equals 1.
2. The $\tilde{\mathcal{O}}(\cdot)$ notation suppresses polylogarithmic dependence on T and d.

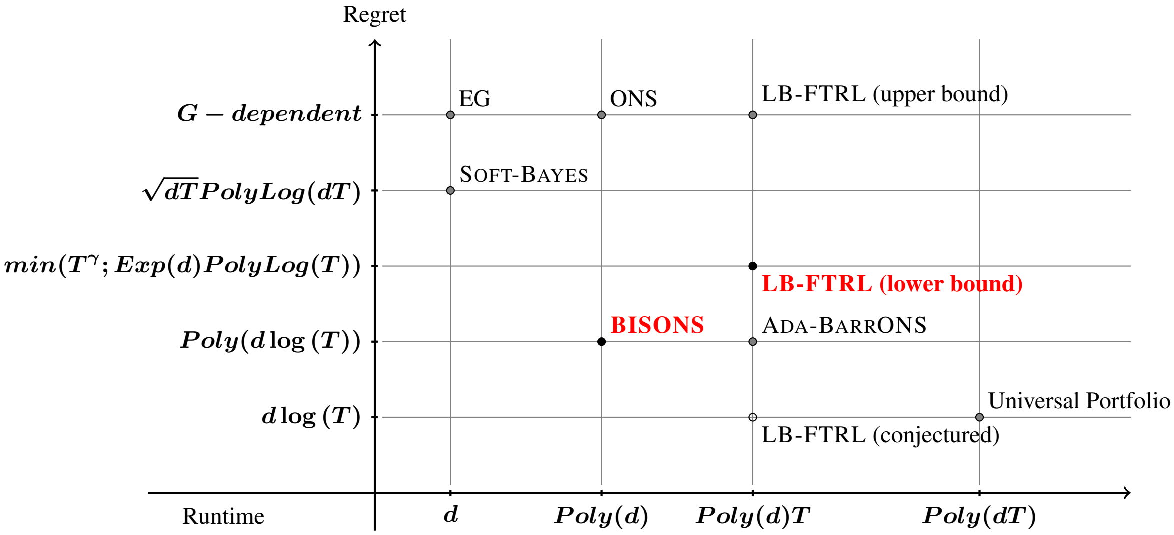

Figure 1: Algorithms for the portfolio problem. Worst-case regret (y-axis) in the $\mathrm{Poly}(d) \ll T$, $T \ll \exp\exp(d)$ regime over per-step computational complexity (x-axis). Our contributions are in red. $0 < \gamma < \frac{1}{2}$ is some universal constant.

lower bound of $\Omega(2^d \log(T) \log\log(T))$ on the regret of the LB-FTRL algorithm. This result effectively removes the LB-FTRL algorithm as a candidate for an optimal trade-off between complexity and regret, since our algorithm obtains superior regret (when $T \le \exp\exp(d)$) at a significantly better run-time and memory complexity.

Related work. The classical online portfolios has a rich literature starting with Cover (1991), who presented the Universal Portfolios algorithm with optimal regret. However, its fastest known implementation (Kalai and Vempala, 2000) requires $\mathcal{O}(T^2(T+d)d^2)$ average per-step computation. Motivated by this inefficiency, early work (Agarwal et al., 2006; Hazan et al., 2007; Hazan and Kale, 2015) develped very efficient second order algorithms – the primary one being Online Newton Step (ONS) – for this problem, under the assumption that the returns of any stock are bounded away from 0 on any day. This assumption translates to a bound G on the gradient of the loss function. ONS obtains $\mathcal{O}(Gd\log(T))$ regret at a per-step computational complexity of $\tilde{\mathcal{O}}(d^3)$. Simpler first order methods based on online gradient descent (Zinkevich, 2003) or multiplicative weights update (Helmbold et al., 1998) can also be applied to the problem, obtaining regret bounds of $\mathcal{O}(G\sqrt{T\log(d)})$ and $\mathcal{O}(G\sqrt{T})$ respectively, at a per step complexity of $\tilde{\mathcal{O}}(d)$.

Since Cover's original work did not have a dependence on G, recent work has focused on overcoming the dependency on G via both first and second order methods. The SOFT-BAYES algorithm (Orseau et al., 2017) is a first order method that obtains $\mathcal{O}(\sqrt{dT\log(d)})$ regret, while preserving linear run-time in d. ADA-BARRONS (Luo et al., 2018) is a second order method based on ONS and achieves $\mathcal{O}(d^2\log^4(T))$ regret. However, it requires computing the solution of log-barrier FTRL at any point, which increases its per-step complexity to $\tilde{\mathcal{O}}(d^{2.5}T)$.

Concurrently Mhammedi and Rakhlin (2022) also improved ADA-BARRONS and obtained $\mathcal{O}(d^2\log^5(T))$ regret in $d^2\log(T)$ memory and $d^3\log(T)$ computational complexity respectively.

The tradeoff between regret and computational complexity described above is plotted schematically in Figure 1. Characterizing the Pareto frontier of this tradeoff has been a subject of study over two decades. In particular, special attention has been given to the log-barrier FTRL algorithm (Agarwal and Hazan, 2005), which obtains a regret of $\mathcal{O}(\min\{G^2 d\log(T), d\log^d(T)\})$, but has been conjectured to obtain the optimal $\mathcal{O}(d\log(T))$ regret by Van Erven et al. (2020).

The online learning of quantum states problem has a shorter history, being introduced by Aaronson et al. (2018). While the log loss version of the problem hasn't been studied before, it is easy to see that the log loss is 1-mixable (Vovk, 1995), and hence Vovk's Aggregating Algorithm can be applied to the problem to obtain an algorithm with $\mathcal{O}(d^2\log(T))$ regret – in fact, this algorithm exactly coincides with Cover's Universal Portfolios algorithm in the online portfolio setting. Implementing this algorithm however is computationally rather inefficient.

Notation. For a natural number d we define $[d] := \{1, 2, \ldots, n\}$, and $\Delta([d])$ to be the set of distributions over $[d]$, seen as vectors in $\mathbb{R}^d$. We denote the set of $d \times d$ Hermitian matrices by $\mathcal{H}^d$. We denote the set of $d \times d$ positive semi-definite Hermitian matrices by $\mathcal{H}^d_+$. Through the paper $\|\cdot\|_p$ denotes the ℓ_p norm. Given a vector v and a positive semi-definite matrix M, we define the semi-norm $\|v\|_M := \sqrt{\mathrm{Tr}(v^* M v)}$. Given two Hermitian matrices X, Y we define the standard inner product (which is always a real number) between them as $\langle X, Y\rangle := \mathrm{Tr}(X^*Y) = \mathrm{Tr}(XY)$. We define additional notation required for the analysis of the quantum learning problem in the Appendix D.

We use the acronyms PSD for positive semi-definite Hermitian matrices and PD for positive definite Hermitian matrices. In general, throughout the paper we denote matrices with capital letters and vectors by small letters. When denoting functions, capital letters are reserved for functions that are defined as sums of functions.

2. Problem setting

Online Optimal Portfolio: The agent interacts with the environment in finite time-steps $t = 1, \ldots, T$. At any time-step, the agent picks a portfolio distribution $x_t \in \mathcal{A} = \Delta([d])$, observes a non-negative returns vector $r_t \in \mathbb{R}^d_+$ and suffers the log loss

$$f_t(x_t) = f(x_t; r_t) := -\log(\langle x_t, r_t\rangle)\,.$$

Since multiplicative scaling of r_t shifts the loss by a constant independent of x_t, the regret is unchanged if we scale r_t so that it lies in $\mathcal{A}$. The goal of the agent is to minimize its regret, defined as the cumulative loss compared to the best static action in hindsight.

$$\mathrm{Reg} = \max_{u\in\mathcal{A}} \mathrm{Reg}(u) = \max_{u\in\mathcal{A}} \sum_{t=1}^{T} \left(f_t(x_t) - f_t(u)\right)\,. \tag{1}$$

Quantum Learning with Log Loss: This problem generalizes the online optimal portfolios problem as follows. The agent's action set is $\mathcal{A} := \{X | X \in \mathcal{H}^d_+, \mathrm{Tr}(X) = 1\}$. The agent at every round picks a PSD Hermitian matrix $X_t \in \mathcal{A}$, observes a PSD loss matrix R_t. As in the portfolios case, the regret is invariant to multiplicative scalings of R_t, so instead of assuming that its eigenvalues lie in $[0, 1]$ as mentioned in the introduction, we may equivalently assume that $R_t \in \mathcal{A}$. The agent then suffers the log loss

$$f_t(X_t) = f(X_t; R_t) := -\log(\langle X_t, R_t\rangle)\,.$$

The task of the agent is to mimimize regret defined analogously to (1). In the Appendix B.1, we show that the above problem formulation captures problem of online learning of quantum states with log loss as described in Aaronson et al. (2018).

3. Algorithm

Algorithm 1: BISONS

input: T, B, η, β, ε_x, ε_u.

initialize: $\forall e \in \mathbb{N}: \ p_0^e = d\mathbf{1}, G_0^e(\cdot) = \hat{F}_0^e(\cdot) = \eta^{-1}R(\cdot), x_1^e = u_1^e = \arg\min_{x\in\mathcal{A}} G_0^e(x)$.

$e \leftarrow 1, \tau \leftarrow 1$

for $t = 1, \ldots$ **do**

  $f_t \leftarrow$ receive from playing $x_t \leftarrow x_\tau^e$.

  $\hat{f}_\tau^e = \hat{f}_t \leftarrow$ construct according to (2).

  $\hat{F}_\tau^e \leftarrow \hat{F}_{\tau-1}^e + \hat{f}_\tau^e$

  $G_\tau^e \leftarrow G_{\tau-1}^e + g_\tau^e$, where $g_\tau^e(x) := \hat{f}_\tau^e(x) - \langle x, p_\tau^e - p_{\tau-1}^e\rangle B$

  $x_{\tau+1}^e \leftarrow$ APPROX-SOLVE$_x(G_\tau^e, x_\tau^e)$, $u_{\tau+1}^e \leftarrow$ APPROX-SOLVE$_u(\hat{F}_\tau^e, u_\tau^e)$

  $\forall i \in [d]: \ p_{\tau+1,i}^e = \max\{p_{\tau,i}^e, {x_{\tau+1,i}^e}^{-1}\}$

  if $\exists i: \ (2(1+6\eta)\beta)u_{\tau+1,i}^e \geq (p_{\tau+1,i}^e)^{-1}$ **then**

    $e \leftarrow e+1, \tau \leftarrow 1$ `// Reset the algorithm`

  else

    $\tau \leftarrow \tau + 1$

  end

end

In this section, we present our main algorithm BISONS (Algorithm 1). The algorithm is inspired by the algorithm ADA-BARRONS proposed by Luo et al. (2018), but improves the regret bound obtained by Luo et al. (2018) by a factor of $\log^2(T)$, while simultaneously and more importantly improving the run-time by factors polynomial in T. BISONS is the first algorithm with constant per-step computational complexity that obtains polylogarithmic regret in the portfolio problem.

The algorithm operates in *epochs*, where each epoch ends when either the global time reaches T or when a certain reset condition (detailed below) is met. We call an epoch *completed* if it ends by reset, which sets the internal time τ of the algorithm back to 1 and lets the algorithm forget all history. Thus, we keep only one copy of all parameters in memory and reset them to the initial values when the epoch is completed.

Let $\mathcal{T}_1 \ldots \mathcal{T}_E \in [1, T]$ denote the timesteps following a restart trigger event. By convention we set $\mathcal{T}_0 = 1$ and $\mathcal{T}_{E+1} := T + 1$. We define an *epoch* $\{\mathcal{E}_i\}$ of the algorithm as the period between successive resets of the algorithm, i.e. $\mathcal{E}_i := [\mathcal{T}_i, \mathcal{T}_{i+1} - 1]$. Note that by definition there is no restriction over the length of these epochs and they can be of variable lengths.

On a high level, BISONS works by approximating at every step, the true loss function $f_t(x)$ by a quadratic surrogate loss

$$\hat{f}_t(x) := f_t(x_t) + \langle x - x_t, \nabla f_t(x_t)\rangle + \frac{\beta}{2}\langle x - x_t, \nabla f_t(x_t)\rangle^2 \,, \tag{2}$$

where $\beta \leq 1$ is an input parameter to the algorithm. Let e, τ be the epoch and internal time of the algorithm at time t, then we define $x_t = x^e_\tau$ and $\hat{f}^e_\tau = \hat{f}_t$. For reasons that become clear in section 4, BISONS further augments the above surrogate loss with a linear bias term, defined at every internal step τ as

$$g^e_\tau(x) := \hat{f}^e_\tau(x) - \langle x, p^e_\tau - p^e_{\tau-1}\rangle B, \tag{3}$$

where $\{p^e_\tau \in \mathbb{R}^d\}$ is an auxiliary sequence maintained by the algorithm and B is a bias scaling factor which is a parameter input to the algorithm. To produce the output x^e_τ BISONS runs approximated FTRL over the biased surrogate losses. Let the FTRL solution be

$$x^{\star e}_\tau := \arg\min_{x\in\mathcal{A}} G^e_{\tau-1}(x) = \arg\min_{x\in\mathcal{A}} \sum_{s=1}^{\tau-1} g^e_s(x) + \eta^{-1}R(x), \tag{4}$$

where η is a learning rate parameter and $R(x) := -\sum_{i=1}^d \log(x_i)$ is the log-barrier regularization. Then approximated FTRL invokes a solver $x_\tau \leftarrow \text{APPROX-SOLVE}(G_{\tau-1}, x_{\tau-1}, \varepsilon_x)$, which outputs an approximated solution $x^e_\tau \approx x^{\star e}_\tau$. The algorithm further maintains a reference solution by running approximated FTRL over the surrogate losses without bias $u^e_\tau \leftarrow \text{APPROX-SOLVE}(\hat{F}^e_{\tau-1}, u^e_{\tau-1}, \varepsilon_u)$, which is close to the FTRL solution

$$u^{\star e}_\tau := \arg\min_{x\in\mathcal{A}} \hat{F}^e_{\tau-1}(x) = \arg\min_{x\in\mathcal{A}} \sum_{s=1}^{\tau-1} \hat{f}^e_s(x) + \eta^{-1}R(x). \tag{5}$$

Further, the asset dependent bias p is updated according to

$$\forall i \in [d]:\; p^e_{\tau,i} = \max\{p^e_{\tau-1,i}, {x^e_{\tau,i}}^{-1}\}\,. \tag{6}$$

Finally, the algorithm is reset (i.e. the bias vector p is reset and all previous losses are discarded) whenever

$$\exists i \in [d]:\; u^e_{\tau+1,i} > \frac{1}{2(1+6\eta)\beta}(p^e_{\tau+1,i})^{-1}.$$

We require the following condition for the solver.

Assumption 1 *We assume* APPROX-SOLVE$_x$, APPROX-SOLVE$_u$ *satisfy at any time* τ*:*

$$\|\nabla G^e_\tau(x_{\tau+1})\|_{[\nabla^2_\Pi G^e_\tau(x_{\tau+1})]^{-1}} \leq \min\{6\eta, 6\sqrt{\eta}\,\|\nabla G^e_\tau(x_\tau)\|_{[\nabla^2_\Pi G^e_\tau(x_\tau)]^{-1}}\}$$
$$\left\|\nabla \hat{F}^e_\tau(x_{\tau+1})\right\|_{[\nabla^2_\Pi \hat{F}^e_\tau(x_{\tau+1})]^{-1}} \leq \frac{1}{8\sqrt{\eta}}\,,$$

where $\nabla^2_\Pi f(x) = \nabla^2 f(x) - \frac{\nabla^2 f(x)\mathbf{1}_d\mathbf{1}_d^\top \nabla^2 f(x)}{\mathbf{1}_d^\top \nabla^2 f(x)\mathbf{1}_d}$[3].

3. $\nabla^2_\Pi f$ can we seen as the projected Hessian over the affine subspace of the actions. We formally define it in the Appendix

The following theorem and corollary capture our main regret bound for BISONS. We show that the total regret in any completed epoch is always non-positive and the total regret in the last uncompleted epoch is bounded. Summing the regrets over individual epochs (which is only an over-estimation of the true regret) gives the final result.

Theorem 1 *Assuming*[4] $T \geq 110d^2$, *the solvers satisfy assumption 1, setting the input parameters as* $B = \frac{264}{5} d \log(T)$, $\eta = \frac{1}{4B}, \beta = \frac{11}{7B}$, *we have that the regret of* BISONS *over a completed (i.e. end triggered by the reset condition) epoch against any comparator* $u : \min_i u_i \geq T^{-1}$ *is non-positive. Further, for the epoch that runs until the end of time* T, *the regret is bounded by* $\mathcal{O}(d^2 \log^2(T))$.

The proof is given in Appendix E, a sketch is provided at the end of Section 4. The following corollary is immediate:

Corollary 2 *Assuming* $T \geq 110d^2$, *the total regret of* BISONS *with parameters from Theorem 1 is bounded by* $\mathcal{O}(d^2 \log^2(T))$.

Runtime: The following lemma shows that given the parameter tuning of Theorem 1, a single damped Newton step update (defined in Lemma 15) is sufficient to solve the minimization problem to the accuracy required by Assumption 1. Hence the per-step runtime is upper bounded by $\mathcal{O}(d^3)$ via vanilla matrix inversion (see Appendix F for further details and the proof of the lemma).

Lemma 3 *For all* e, τ *when executing Algorithm 1 with the parameter tuning of Theorem 1 it is sufficient to apply one damped Newton step (defined in Lemma 15) in the* APPROX-SOLVE$_x$ *and* APPROX-SOLVE$_u$ *subroutines to satisfy Assumption 1.*

3.1. Extension to Quantum Learning

In this section, we describe the SCHRÖDINGER'S-BISONS algorithm (formally defined in the appendix as Algorithm 3) for the quantum learning problem. SCHRÖDINGER'S-BISONS follows the same structure as BISONS, and uses the same choice of surrogate function $\hat{f}_t$, point played X_t, and comparator U_t as in online optimal portfolio which are still well defined by (2), (4) and (5) respectively.

We highlight the main differences from the online optimal portfolio in this section. The main differences between the two cases firstly is that the regularizer R used is the log-det-barrier, which reduces to the log-barrier for diagonal matrices: $R(X) = -\log\det(X)$. Secondly, and the primary non-trivial step in the generalization, is the appropriate definition of the biases P_t and the reset condition. Analogous to Algorithm 1, the reset condition is generalised to $U_t^e \not\preceq \frac{1}{2(1+6\eta)\beta}[P_t^e]^{-1}$, for some biases P_t^e ensuring $P_t^e \succeq [X_s^e]^{-1}$ for all s, t in the same epoch with $s \leq t$. This ensures that within any epoch $\hat{f}_t^e$ stays a valid lower bound for the comparator U_τ^e for that epoch. This property is summarized as Lemma 31 in the appendix.

The main hurdle for extending our results to the quantum setting is to find a suitable bias rule P_τ^e that generalises (6). The goal is to construct P_τ^e that satisfies $P_{\tau+1}^e \succeq P_\tau^e$ and $P_\tau^e \succeq [X_\tau^e]^{-1}$. Unlike in the online optimal portfolio case, there is no canonical "smallest" P_τ^e with that property in general. Instead we choose to look for a choice satisfying these constraints that suffers a small

4. Without loss of generality, we can fill up missing time-steps with $r_t = \mathbf{1}_d/d$, which result in constant losses.

cost of bias [5] $\sum_{t=1}^{\tau}\langle X_\tau^e, P_\tau^e - P_{\tau-1}^e\rangle$. This objective, which can be characterized via semi-definite programming duality, leads to an optimal choice given by

$$P_{\tau+1}^e = P_\tau^e + [X_{\tau+1}^e]^{-\frac{1}{2}}\left(\mathbf{I}_d - [X_{\tau+1}^e]^{\frac{1}{2}}P_\tau^e[X_{\tau+1}^e]^{\frac{1}{2}}\right)_+ [X_{\tau+1}^e]^{-\frac{1}{2}}\,, \tag{7}$$

where $(\cdot)_+$ is the operator that sets all negative eigenvalues to 0, i.e. if M is a Hermitian matrix with eigendecomposition $M = U^*PU + V^*NV$, where P and N are diagonal matrices with the non-negative and negative eigenvalues respectively, and the columns of U and V are the corresponding eigenvectors, then $M_+ = U^*PU$.

Remark 4 *For diagonal matrices,* (7) *picks* $P_\tau^e(i,i) = \max\{[X_\tau^e]^{-1}(i,i), P_{\tau-1}^e(i,i)\}$ *and is hence a strict generalization of* (6)*.*

Surprisingly, we show in the appendix that the cost of bias remains $\mathcal{O}(d\log(T)B)$, so we do not pay anything for this generalization. We note that relying on the "negative regret via linear bias" technique used here is crucial towards obtaining this generalization. It is not clear how to use the "negative regret by increasing learning rate" approach used in ADA-BARRONS here. We now state the theorem governing the regret for SCHRÖDINGER'S-BISONS.

Theorem 5 *Assuming* $T \geq 110d^2$*, setting* $B = \frac{264}{5}d^2\log(T)$*,* $\eta = \frac{1}{4B}, \beta = \frac{11d}{7B}$ *the regret of* SCHRÖDINGER'S-BISONS *over a single epoch against any comparator* $U \succeq T^{-1}\mathbf{I}_d$ *is non-positive if the end is triggered by the reset condition. Otherwise, if the algorithm runs until the end of time* T*, then the regret is bounded by* $\mathcal{O}(d^3\log^2(T))$*.*

Theorem 5 can be used to prove the following regret bound for SCHRÖDINGER'S-BISONS yields the following corollary analogous to Corollary 2. Missing proofs are in Appendix E.

Corollary 6 *For* $T \geq 110d^2$*, the regret of* SCHRÖDINGER'S-BISONS *is bounded by* $\mathcal{O}(d^3\log^2(T))$.

4. Overview of the Analysis

Intuition for the regret bound. Using quadratic surrogate losses instead of the true losses is a standard technique for improving computation complexity while preserving logarithmic regret (see the Online Newton Step (ONS) method from Hazan et al. (2007)). We use the same quadratic surrogate $\hat{f}_t$ as the ONS method (with a different choice of β). Such analyses including ONS often require that the surrogate is a lower bound for the function value of the comparator u, i.e. $\hat{f}_t(u) \leq f_t(u)$ at all time-steps. Since u is unknown, this is typically enforced by ensuring lower boundedness over the entire domain. However in the case of optimal portfolio, a uniform lower bound requires β to scale inversely with the largest observed gradient G, a quantity we wish to avoid in our bound.

Luo et al. (2018) observe that for any t, $\hat{f}_t(u) > f_t(u)$ only if there exists i such that $u_i = \Omega(\frac{x_{t,i}}{\beta})$. Intuitively this condition is triggered when the stock i underperformed up to time t, thereby receiving a low weight from the algorithm, but later on recovers overproportionally. To counter this case, our algorithm biases stocks to give them more weight according to the poorest performance they experienced. The bias term we introduce in our algorithm ensures a negative contribution to appear in the regret analysis. This quantity is carefully tuned such that, if a reset happens, the regret

5. See Section 4 for an explanation of what cost of bias means and how it shows up in the analysis.

for this phase is non-positive. To demonstrate how the negative regret contribution appears, consider the following decomposition of the surrogate losses:

$$\sum_{t=1}^{\tau}(\hat{f}_t(x_t) - \hat{f}_t(u)) = \sum_{t=1}^{\tau}(g_t(x_t) - g_t(u) + \langle x_t - u, p_t - p_{t-1}\rangle B)$$
$$= \underbrace{\mathrm{Reg}_g(u)}_{\text{FTRL regret bound}} + \underbrace{\sum_{t=1}^{\tau}\langle x_t, p_t - p_{t-1}\rangle B}_{\text{cost of bias}} \underbrace{-\langle u, p_\tau - p_0\rangle B}_{\text{negative regret}}.$$

The FTRL regret over the sequence of functions g_t is bounded via ONS analysis. Further recall that the bias parameters p_t satisfy for all i, $p_{ti} = \max_{s\le t} x_{ti}^{-1}$. Therefore for all t, i, $p_{ti} - p_{t-1,i} \neq 0$ implies $x_{ti} = p_{ti}^{-1}$. We can now bound the *cost of bias* is any epoch by

$$\sum_{t=1}^{\tau}\langle x_t, p_t - p_{t-1}\rangle B = \sum_{i=1}^{d}\sum_{t=1}^{\tau} p_{ti}^{-1}(p_{ti} - p_{t-1,i})B \le \sum_{i=1}^{d}\log(p_{\tau i}/d)B\,.$$

We show in our analysis that $p_{ti} \le T^2$ at all time-steps, so this term is bounded by $\mathcal{O}(d\log(T)B)$. If a reset is triggered at timestep τ, then by the reset condition we have for the comparator u_τ (maintained by the algorithm), $\exists i \in [d] : u_{\tau i}p_{\tau i} = \Omega(\beta^{-1})$. Hence the negative regret is of order $\Omega(\frac{B}{\beta})$, which is, given the right tuning, significantly larger than the cost of bias. We argued the above for the comparator u_τ maintained by the algorithm, which is the FTRL solution of the quadratic surrogate losses $\hat{f}_t$. This choice of comparator is the core reason behind our runtime improvement. We now explain why this works.

Improving the run-time. The key to our improved runtime complexity is using the FTRL solution over the surrogate losses as comparator for the reset condition. This computation is as costly as x_t, which can be done in $\mathcal{O}(d^{2.5})$ arithmetic operations, in contrast to $\mathcal{O}(dT)$ required by previous algorithms with optimal regret, e.g. ADA-BARRONS (Luo et al., 2018). We first setup some auxiliary notation to simplify our argument. Let $\ell_t(x) = \langle x, r_t\rangle$ be the linear reward at time t, then we can rewrite $f_t = h \circ \ell_t$ and $\hat{f}_t = \hat{h}_t \circ \ell_t$, where $y_t := \ell_t(x_t)$ and $h(x), \hat{h}_t(x) : \mathbb{R}_+ \to \mathbb{R}$ are functions defined as

$$h(x) := -\log(x)\,, \qquad \hat{h}_t(x) := h(y_t) + (x - y_t)h'(y_t) + \frac{\beta}{2}(x - y_t)^2 h'(y_t)^2\,.$$

Note that both $h, \hat{h}_t$ are convex functions. We now define an additional function $\underline{\hat{f}_t} = \underline{\hat{h}_t} \circ \ell_t$, with $\underline{\hat{h}_t}(x) = \hat{h}_t(x)$ if $x \le \beta^{-1}y_t$, and $\hat{h}_t(\beta^{-1}y_t) + (x - \beta^{-1}y_t)\hat{h}_t'(\beta^{-1}y_t)$ otherwise. Geometrically $\underline{\hat{h}_t}$ coincides with $\hat{h}_t$ for x up to $\beta^{-1}y_t$ and follows its linear extension at $x = \beta^{-1}y_t$ afterwards (see Figure 2). From the convexity of $\hat{h}_t$, it follows that both $\underline{\hat{h}_t}$ and $\underline{\hat{f}_t}$ are convex. Furthermore as shown by the following lemma, it holds that $\underline{\hat{h}_t}$ is a proper lower approximation of h and therefore $\underline{\hat{f}_t}$ is a proper lower approximation of f.

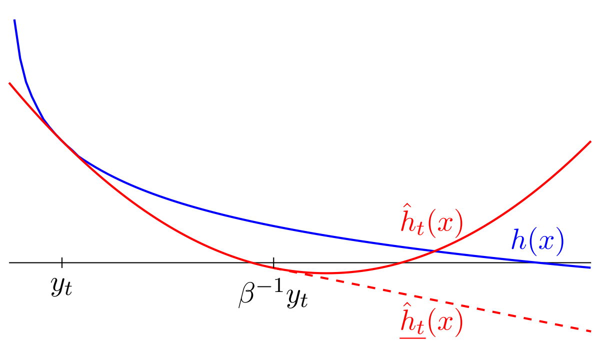

Figure 2: Surrogate losses

Lemma 7 *For all $x \in (0, \infty)$: $\underline{\hat{h}_t}(x) \le h(x)$, where equality holds for $x = y_t$.*

The proof can be found in Appendix E. We have introduced the function $\underline{\hat{f}_t}$ merely as a tool for the analysis. An important invariant of our algorithm that our reset condition ensures is:

Lemma 8 *Let $\eta \le \min\{\frac{1}{4B}, \frac{\beta}{4}, \frac{1}{63}\}$. Consider any epoch e with the reset points $\mathcal{T}_{e-1} < \mathcal{T}_e \le T$. Let L represent the length of the epoch, i.e. $L = \mathcal{T}_e - \mathcal{T}_{e-1}$, we have that, it holds that*

$$\min_{x \in \mathcal{A}} \sum_{\tau=1}^{L} \underline{\hat{f}^e_\tau}(x) + \eta^{-1} R(x) = \sum_{\tau=1}^{L} \hat{f}^e_t(u_{\tau+1}) + \eta^{-1} R(u_{\tau+1}) \,.$$

While we defer the proof to Appendix E, the high level idea is that the reset condition ensures that $\ell_t(u_{\tau+1}) \le \beta^{-1} y_s$ for all $s \le \tau$. That means that the LHS is equal to the RHS around $u_{\tau+1}$. Since $u_{\tau+1}$ by definition is the minimizer of the RHS (which is a strictly convex function), hence it is a local minimizer of the LHS, and thereby due to convexity, also a global minimizer. We are now ready to provide a full proof sketch for Theorem 1.

Proof sketch of Theorem 1. Let τ denote the last time-step of any particular epoch. Then

$$\begin{aligned}
\mathrm{Reg}(u) &= \sum_{t=1}^{\tau} (f_t(x_t) - f_t(u)) \le \sum_{t=1}^{\tau} (\underline{\hat{f}_t}(x_t) - \underline{\hat{f}_t}(u)) && \text{(by Lemma 7)} \\
&\le \max_{u' \in \mathcal{A}} \left(\sum_{t=1}^{\tau} (\underline{\hat{f}_t}(x_t) - \underline{\hat{f}_t}(u')) - \eta^{-1} R(u') + \eta^{-1} R(u) \right) \\
&= \sum_{t=1}^{\tau} (\hat{f}_t(x_t) - \hat{f}_t(u_{\tau+1})) - \eta^{-1} R(u_{\tau+1}) + \eta^{-1} R(u) && \text{(by Lemma 8)} \\
&= \mathrm{Reg}_g(u_{\tau+1}) - \eta^{-1} R(u_{\tau+1}) + \sum_{t=1}^{\tau} \langle x_t - u_{\tau+1}, p_t - p_{t-1} \rangle B + \eta^{-1} R(u) \,.
\end{aligned}$$

We show in the detailed proof that the FTRL regret over g is bounded by $\mathcal{O}(\frac{d}{\beta} \log(T))$ and the regularizer is bounded by $\mathcal{O}(\frac{d}{\eta} \log(T))$ due to the constraint on u. As discussed before, the cost of bias is bounded by $\mathcal{O}(d \log(T) B)$ and the negative regret in case a reset is triggered is of order $\Omega(\frac{B}{\beta})$. Set $\beta = \Theta(\eta) = \Theta(\frac{1}{B})$, then the regret is bounded by

$$\mathrm{Reg}(u) = \mathcal{O}(d \log(T) B) - \Omega(B^2) \mathbb{I}\{\text{reset triggered}\} \,.$$

Finally tuning $B = \Theta(d \log(T))$ completes the proof. ■

Comparison with ADA-BARRONS (Luo et al., 2018) ADA-BARRONS uses the same surrogate loss as us, but computes x_t via online mirror descent (OMD) updates with increasing learning rate. This technique is closely related to using linear biases (see Foster et al. (2020) for a detailed discussion), however as we show via our application to the quantum learning problem (See Section 3.1), the latter is more flexible and additionally saves a $\log(T)$ factor in the regret. ADA-BARRONS does not use a fixed β but instead doubles the parameter β_e with every reset. They ensure bounded regret by tuning the negative regret of phase e, such that it cancels the Reg_g term of the next phase $e+1$. Additionally, they show that the total number of epochs is bounded by $\log(T)$. We go a step further and not only cancel the Reg_g term, but all positive regret contributions. This allows us to use a fixed β and saves another $\log(T)$ factor in the regret. Finally, our algorithm uses the FTRL solution over surrogate losses instead of the FTRL solution over the true losses for the comparator u_t as run by ADA-BARRONS. This is made possible via the introduction of the auxiliary functions $\underline{f_t}$ combined with Lemma 8 and yields the improvement in computational complexity.

5. Lower bound for FTRL

In this section, we disprove a COLT 2020 conjecture (Van Erven et al., 2020) regarding FTRL for the online portfolio selection problem. Throughout this section, we consider FTRL with regularizer $R(x) = -\sum_{i=1}^{d} \log(x_i)$, simply referred to as LB-FTRL. In round t, this algorithm plays $x_t := \arg\min_{x\in\mathcal{A}} F_t(x)$, where $F_t(t) := \eta^{-1}R(x)+\sum_{\tau=1}^{t-1} f_t(x)$ and $\eta > 0$ is a constant hyperparameter. This is in some sense a natural choice, since the adversary can "force" the player to operate with this regularization by picking $r_i = \boldsymbol{e}_i$ for $i \in [d]$. Indeed Van Erven et al. (2020) conjectured that FTRL obtains the optimal bound of $\mathcal{O}(d\log(T))$, while we prove an exponentially worse lower bound of $\Omega(2^d \log(T)\log\log(T))$. Our main theorem is the following (all missing proofs appear in Appendix G):

Theorem 9 *The worst-case regret of* LB-FTRL *for any* $T > Poly(2^d)$ *is* $\Omega(2^d \log(T)\log\log(T))$.

Remark. This lower bound extends easily to the quantum version of LB-FTRL which uses the log-det regularizer via the observation that when all the loss matrices R_t are diagonal, log-det regularized LB-FTRL reduces to vanilla (log barrier regularized) LB-FTRL.

Lower bound proof sketch. First, we note that the action set $\Delta([d])$ lies in a $(d-1)$-dimensional subspace of $\mathbb{R}^d$. For technical reasons, it will be convenient to work with a full dimensional action set with non-zero volume. Hence, we define the projection operator $\Pi : \mathbb{R}^d \to \mathbb{R}^{d-1}$ with kernel $\boldsymbol{c} = \frac{1}{d}\mathbf{1}_d$ and Π^{-1} its inverse mapping into $\mathcal{A}$.[6] Thus $\mathcal{A}$ gets mapped to $\Pi\mathcal{A}$, which has non-zero volume in $\mathbb{R}^{d-1}$. In a slight overload of notation, we consider $f^{\Pi}(\tilde{x}; r)$ as a function with argument $\tilde{x} \in \Pi\mathcal{A}$ by the identity

$$f^{\Pi}(\tilde{x}; r) = f(\Pi^{-1}\tilde{x}; r) = -\log(\langle \Pi^{-1}\tilde{x}, r\rangle) = -\log(1/d + \langle \tilde{x}, \Pi r\rangle)\,,$$

and use $\nabla_\Pi f_t(x) = \nabla f^{\Pi}(\Pi x; r_t)$, $\nabla^2_\Pi f_t(x) = \nabla^2 f^{\Pi}(\Pi x; r_t)$ as shorthand notation for the gradient and Hessians with respect to the above definition of f_Π. We define $\nabla_\Pi F_t(x)$ and $\nabla^2_\Pi F_t(x)$ analogously.

The lower bound rests on the following key lemma, which shows that the regret of LB-FTRL is lower bounded by a certain *stability* quantity which also appears in the *upper bound* for FTRL in the standard analysis; so the stability controls the regret tightly. This is, to the best of our knowledge, a novel idea and is crucial in showing that LB-FTRL does not obtain $\tilde{\mathcal{O}}(d\log(T))$ regret in the portfolio problem.

6. Let U be a $d \times (d-1)$ matrix whose columns form an orthonormal basis for the subspace orthogonal to $\boldsymbol{c}$. Then Π can be defined as $\Pi x = U^* x$, and Π^{-1} as $\Pi^{-1} v = Uv + \boldsymbol{c}$.

Lemma 10 *The regret of* LB-FTRL *is lower bounded as follows:*

$$\mathrm{Reg} = \Omega\left(\sum_{t=1}^{T} \|\nabla_\Pi f_t(x_t)\|^2_{(\nabla^2_\Pi F_t(x_t))^{-1}}\right).$$

We now give a high level intuition of why a lower bound of $\mathcal{T}\log(T)$ is possible. The extra $\log\log(T)$ factor requires a careful layering construction that is deferred to the appendix. The lower bound relies on certain "target" portfolio vectors and associated returns vectors that satisfy a particular admissibility property:

Definition 11 *For some integer $\mathcal{T} > 0$, a sequence of target portfolio vectors $\boldsymbol{t}_1, \dots, \boldsymbol{t}_{\mathcal{T}} \in \Delta([d])$ and associated returns vectors $\boldsymbol{o}_1, \dots, \boldsymbol{o}_{\mathcal{T}} \in \Delta([d])$, are called* admissible *if $\forall j < i : \langle \boldsymbol{t}_i, \boldsymbol{o}_j\rangle = \Omega(1/\mathit{Poly}(d))$, and $\forall i : \langle \boldsymbol{t}_i, \boldsymbol{o}_i\rangle = 0$.*

Given admissible sequences of target portfolio and returns vectors, we can now use them to define adversarial sequences of returns for the LB-FTRL learner. This is done in Algorithm 4 (Appendix G), whose simplified form appears in Algorithm 2 below. Here, we use the notation $\boldsymbol{t}'_i = (1 - T^{-\alpha})\boldsymbol{t}_i + T^{-\alpha}\boldsymbol{c}$ and $\boldsymbol{o}'_i = (1 - T^{-\alpha})\boldsymbol{o}_i + T^{-\alpha}\boldsymbol{c}$.

Algorithm 2: Sequence for large regret (simplified version).

```
Input: (t_i, o_i)_{i=1}^𝒯, α = 1/8, T
for i = 1, 2, ..., 𝒯 do
    for k = 1, ..., T^α do
        while x_t ≠ t'_i do
            r_t ← move-to-x(t'_i; F_{t-1})
            t ← t + 1
        end
        r_t ← o'_i
        t ← t + 1
    end
end
Function move-to-x(x; F):
    g ← Π∇F(x)
    g ← min{T^{-1/2} / ||g||_2 , 1/(d max{1 − ⟨g, Πx⟩, 0})} g
    return: Π^{-1} g
```

The following key lemma gives a lower bound on the regret of LB-FTRL when Algorithm 4 is used to generate the returns vectors:

Lemma 12 *Suppose that for some integer $\mathcal{T} > 0$, $\boldsymbol{t}_1, \dots, \boldsymbol{t}_{\mathcal{T}} \in \Delta([d])$ and $\boldsymbol{o}_1, \dots, \boldsymbol{o}_{\mathcal{T}} \in \Delta([d])$ are admissible sequences of target portfolio and returns vectors. Then there exists $T_0 = \mathit{Poly}(\mathcal{T}, d)$, such that for any $T > T_0$ the regret of* LB-FTRL *against the sequence of reward vectors r_t generated by Algorithm 4 (Appendix G) using these sequences is lower bounded by*

$$\mathrm{Reg} = \Omega(\mathcal{T}\log(T)\log\log(T)).$$

The main idea of Algorithm 4 is to force the LB-FTRL learner to sequentially visit each of the points $(\boldsymbol{t}_i)_{i=1}^{\mathcal{T}}$ for T^α steps ($0 < \alpha < 1$ is some fixed parameter), and receive the return $\boldsymbol{o}_i$ at point $\boldsymbol{t}_i$. Since $\boldsymbol{t}_i$ may be on the boundary of $\Delta([d])$, which the LB-FTRL learner cannot reach exactly, we refer to visiting $\boldsymbol{t}_i$ if the learner plays $\boldsymbol{t}'_i = (1 - T^{-\alpha})\boldsymbol{t}_i + T^{-\alpha}\boldsymbol{c}$, which is the target pulled towards the center by $T^{-\alpha}$. Let us first assume that it is possible to force the LB-FTRL learner to visit the points $\boldsymbol{t}_i$ as stated, and that we only need to account for the times that the learner plays the $\boldsymbol{t}_i$ points in the Hessian $\nabla^2_\Pi F_t(x_t)$ in Lemma 10. We can now analyze the regret by simplifying the stability terms in Lemma 10 via the following bound:

$$\|\nabla_\Pi f_t(x_t)\|^2_{(\nabla^2_\Pi F_t(x_t))^{-1}} \geq \frac{\|\nabla_\Pi f_t(x_t)\|^2}{\mathrm{Tr}(\nabla^2_\Pi F_t(x_t))}.$$

During the T^α times we visit $\boldsymbol{t}_i$ and receive $\boldsymbol{o}_i$, the term $\|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_i)\|^2$ is of order $T^{2\alpha}$ (ignoring dimension dependence), since it scales with $\langle (1 - T^{-\alpha})\boldsymbol{t}_i + T^{-\alpha}\boldsymbol{c}, \boldsymbol{o}_i\rangle^{-2} = T^{2\alpha}\langle \boldsymbol{c}, \boldsymbol{o}_i\rangle^{-2}$. The trace in the denominator (ignoring the regularizer) after the m-th visit of $\boldsymbol{t}_i$, is

$$\sum_{j<i} T^\alpha \left\|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_j)\right\|^2 + m \left\|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_i)\right\|^2 = \mathcal{O}(\mathcal{T}\mathrm{Poly}(d)T^\alpha) + m \left\|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_i)\right\|^2 ,$$

which uses $\langle \boldsymbol{t}'_i, \boldsymbol{o}_j\rangle = \Omega(1/\mathrm{Poly}(d))$ for $i < j$. We can assume that T is large enough that $T^{\alpha/2} = \Omega(\mathrm{Poly}(d))$, so for any $m > T^{\alpha/2}$, the denominator is of order $\mathcal{O}(m \|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_i)\|^2)$. Hence the stability in Lemma 10 is approximated by

$$\sum_{i=1}^{\mathcal{T}} \sum_{m=T^{\alpha/2}}^{T^\alpha} \frac{\|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_i)\|^2}{m \|\nabla_\Pi f(\boldsymbol{t}'_i; \boldsymbol{o}_i)\|^2} = \Omega(\mathcal{T} \log(T)) \,.$$

This shows that the stability is large if the LB-FTRL learner's trajectory can be controlled to visit the $\boldsymbol{t}_i$'s as specified. In fact, this is possible without increasing the trace of the Hessian significantly. At any step where the learner plays $\boldsymbol{t}'_i$ and receives $\boldsymbol{o}_i$, the next LB-FTRL iterate will move away from $\boldsymbol{t}'_i$; we then force the learner to move back to $\boldsymbol{t}'_i$ by interleaving the $\boldsymbol{o}_i$ returns with additional *movement-returns* r_t (see the `move-to-x` subroutine in Algorithm 2), which satisfy $\|\nabla\Pi r_t\| = \mathcal{O}(T^{-\frac{1}{2}})$. Since the contribution to the Hessian is roughly quadratic in $\|\nabla\Pi r_t\|$, the cumulative contribution to the Hessian trace of all movement steps does not exceed $\mathcal{O}(\mathrm{Poly}(d))$, which is negligible in the argument above. Finally, one needs to show that the required number of *movement-returns* is small enough such that the sequence does not exceed T time steps. In our detailed proof, we show that this always holds for $\alpha = \frac{1}{8}$ and sufficiently large T.

Finally, equipped with Lemma 12, we are ready to derive an exponential lower bound for LB-FTRL. All we need to do is construct appropriate admissible sequences of target portfolio and returns vectors. We define the following sequence of target point sets for any $k \in [d-1]$: $\boldsymbol{\mathcal{T}}_k := \left\{\frac{1}{k}x \,\middle|\, x \in \{0,1\}^d, \|x\|_1 = k\right\}$, i.e. the sets where exactly k components of the vector are non-zero, and these are of equal size. Define the combined sequence by adding the sets in increasing order of k, with arbitrary ordering within a set $\boldsymbol{t}_1, \dots, \boldsymbol{t}_{\mathcal{T}} = (\boldsymbol{t} \in \boldsymbol{\mathcal{T}}_1), \dots, (\boldsymbol{t} \in \boldsymbol{\mathcal{T}}_{d-1})$. For each $i \in [\mathcal{T}]$, define the associated returns vector by $\boldsymbol{o}_i := \frac{1}{d - \|\boldsymbol{t}_i\|_0}(\mathbf{1}_d - \|\boldsymbol{t}_i\|_0\, \boldsymbol{t}_i)$, i.e. the complement vector that is non-zero iff $\boldsymbol{t}_i$ is zero, normalized so that it lies in $\Delta([d])$. The following lemma now shows that these sequences are admissible:

Lemma 13 *For all $i < j$, it holds $\langle \boldsymbol{t}_i, \boldsymbol{o}_j \rangle = \Omega(1/d^2)$, as well as $\langle \boldsymbol{t}_i, \boldsymbol{o}_i \rangle = 0$.*

Proof The second equality follows trivially by construction. For the first observe that for any $j < i$, the number of non-zero components in $\boldsymbol{t}_j$ does not exceed the number of non-zero components in $\boldsymbol{t}_i$. That means that if $\boldsymbol{t}_i$ has k non-zero entries, then $\boldsymbol{o}_j$ has at least $d - k$ non-zero entries. Since $\boldsymbol{t}_j \neq \boldsymbol{t}_i$, $\boldsymbol{o}_j \neq \boldsymbol{o}_i$, there is at least one component of non-zero values overlapping. Finally all non-zero components are least of size $\frac{1}{d}$, which completes the proof. ■

Theorem 9 now follows immediately:

Proof [Theorem 9] Combine Lemma 13 with Theorem 12 and observe that the constructed sequence is of length $2^d - 2$. ■

6. Conclusion

We have presented BISONS, the first algorithm with $\tilde{\mathcal{O}}(\text{poly}(d))$ memory and per-step running time that obtains near optimal regret in the optimal portfolio problem without any assumptions on the gradient. Further, we have shown that key techniques in our algorithm BISONS can be adapted to work with the more general setting of quantum learning with log loss as well, at an additional factor of d in the regret.

Further, we showed that previous conjectures about LB-FTRL are wrong and that the worst-case regret of LB-FTRL is at least of order $2^d \log(T) \log\log(T)$. In the natural regime of $T \ll \exp(\exp(d))$ the regret of BISONS outperforms LB-FTRL at a significantly lower run-time and memory complexity. Therefore we practically eliminate LB-FTRL as a candidate for optimal trade-off between regret and computational complexity.

References

Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. In *NeurIPS*, pages 8976–8986, 2018.

Amit Agarwal and Elad Hazan. Efficient algorithms for online game playing and universal portfolio management. *ECCC*, TR06-033, 2005.

Amit Agarwal, Elad Hazan, Satyen Kale, and Robert E Schapire. Algorithms for portfolio management based on the newton method. In *Proceedings of the 23rd international conference on Machine learning*, pages 9–16, 2006.

Thomas M Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Dylan J Foster, Claudio Gentile, Mehryar Mohri, and Julian Zimmert. Adapting to misspecification in contextual bandits. *Advances in Neural Information Processing Systems*, 33, 2020.

Elad Hazan and Satyen Kale. An online portfolio selection algorithm with regret logarithmic in price variation. *Mathematical Finance*, 25(2):288–310, 2015.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

David P Helmbold, Robert E Schapire, Yoram Singer, and Manfred K Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8(4):325–347, 1998.

Are Hjorungnes and David Gesbert. Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing*, 55(6):2740–2746, 2007.

Adam Kalai and Santosh S. Vempala. Efficient algorithms for universal portfolios. In *FOCS*, pages 486–491. IEEE Computer Society, 2000.

Haipeng Luo, Chen-Yu Wei, and Kai Zheng. Efficient online portfolio with logarithmic regret. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 8245–8255, 2018.

Zakaria Mhammedi and Alexander Rakhlin. Damped online newton step for portfolio selection. *arXiv preprint arXiv:2202.07574*, 2022.

Yurii Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, 2018.

Peter J Olver, Chehrzad Shakiban, and Chehrzad Shakiban. *Applied linear algebra*, volume 1. Springer, 2006.

Laurent Orseau, Tor Lattimore, and Shane Legg. Soft-bayes: Prod for mixtures of experts with log-loss. In *International Conference on Algorithmic Learning Theory*, pages 372–399. PMLR, 2017.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

Tim Van Erven, Dirk Van der Hoeven, Wojciech Kotłowski, and Wouter M Koolen. Open problem: Fast and optimal online portfolio selection. In *Conference on Learning Theory*, pages 3864–3869. PMLR, 2020.

V. G. Vovk. A game of prediction with expert advice. In *COLT*, pages 51–60. ACM, 1995.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936, 2003.

Appendix A. Self-concordant functions: Definitions and useful properties

Definition 14 *A convex function $f : \mathcal{K} \to \mathbb{R}$ over $\mathcal{K} \subset \mathbb{R}^n$ is called M self-concordant, if f is in $\mathcal{C}^3$, $f(x_k) \to \infty$ for $x_k \to x \in \partial\mathcal{K}$ and $\forall x \in \mathcal{K}, \forall u \in \mathbb{R}^n$:*

$$|\nabla^3 f(x)[u, u, u]| \leq 2M \left\| u \right\|^3_{\nabla^2 f(x)} .$$

The following two Lemmas are standard results (Nesterov et al., 2018).

Lemma 15 *Let f be an M self-concordant function over $\mathcal{K}$ and $x, y \in \mathcal{K}$ such that*

$$\|x-y\|_{\nabla^2 f(x)} < 1/M\,,$$

then it holds

$$(1 - M\,\|x-y\|_{\nabla^2 f(x)})^2 \nabla^2 f(y) \preceq \nabla^2 f(x) \preceq \frac{1}{(1 - M\,\|x-y\|_{\nabla^2 f(x)})^2} \nabla^2 f(y)\,.$$

Additionally, the damped newton step update

$$x_+ := x - \frac{[\nabla^2 f(x)]^{-1} \nabla f(x)}{1 + M\,\|\nabla f(x)\|_{[\nabla^2 f(x)]^{-1}}}$$

satisfies for any x such that $\|\nabla f(x)\|_{[\nabla^2 f(x)]^{-1}} < 1/M$:

$$\|\nabla f(x_+)\|_{[\nabla^2 f(x_+)]^{-1}} \leq 2M\,\|\nabla f(x)\|^2_{[\nabla^2 f(x)]^{-1}}\,.$$

Lemma 16 *For an M self-concordant function f, the function αf for $\alpha > 0$ is $M/\sqrt{\alpha}$ self-concordant.*

As corollaries, we get the following useful properties.

Lemma 17 *Let f be an M self-concordant function over $\mathcal{K} \subset \mathbb{R}^n$ and $x, y \in \mathcal{K}$. Then it holds*

$$\min_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)} \geq \frac{\max_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)}}{1 + M \max_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)}}\,.$$

Proof If $\min_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)} \geq \frac{1}{M}$, then the lemma follows immediately. Otherwise assume $z_0 = \arg\min_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)}$, $z_1 = \arg\max_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)}$ and $\frac{1}{M} > \|x-y\|_{\nabla^2 f(z_0)} \geq \|z_0 - z_1\|_{\nabla^2 f(z_0)}$. We have by Lemma 17

$$\|z_0 - z_1\|_{\nabla^2 f(z_1)} \leq \frac{\|z_0 - z_1\|_{\nabla^2 f(z_0)}}{1 - M\,\|z_0 - z_1\|_{\nabla^2 f(z_0)}} \leq \frac{\|z_0 - z_1\|_{\nabla^2 f(z_0)}}{1 - M\,\|x-y\|_{\nabla^2 f(z_0)}}\,.$$

Since $z_0 - z_1 = c(x-y)$ for some $c \in \mathbb{R}$, this implies

$$\|x-y\|_{\nabla^2 f(z_1)} \leq \frac{\|x-y\|_{\nabla^2 f(z_0)}}{1 - M\,\|x-y\|_{\nabla^2 f(z_0)}}\,.$$

Rearranging completes the proof. ■

Lemma 18 *For any M-self-concordant function f over $\mathcal{K} \subset \mathbb{R}^n$ and any $x, y \in \mathcal{K}$, it holds that*

$$\min_{z \in [x,y]} \|\nabla f(x) - \nabla f(y)\|_{[\nabla^2 f(z)]^{-1}} = \xi < \frac{1}{M}$$
$$\Rightarrow \max_{z \in [x,y]} \|x-y\|_{\nabla^2 f(z)} \leq \frac{\xi}{1 - \xi M}\,.$$

Proof We prove the statement

$$\min_{z\in[x,y]} \|\nabla f(x) - \nabla f(y)\|_{[\nabla^2 f(z)]^{-1}} \geq \frac{\max_{z\in[x,y]} \|x-y\|_{\nabla^2 f(z)}}{1 + \max_{z\in[x,y]} \|x-y\|_{\nabla^2 f(z)} M}\,,$$

from which the Lemma follows directly by rearranging.

Denote $z(\alpha) = x + \alpha(y-x)$, and $z' = \arg\,\min_{z\in[x,y]} \|\nabla f(x) - \nabla f(y)\|_{[\nabla^2 f(z)]^{-1}}$, then

$$\begin{aligned}\int_0^1 \|x-y\|^2_{\nabla^2 f(z(\alpha))}\, d\alpha = \langle \nabla f(x) - \nabla f(y), x-y\rangle &\leq \|x-y\|_{\nabla f^2(z')} \|\nabla f(x) - \nabla f(y)\|_{[\nabla^2 f(z')]^{-1}} \\ &\leq \max_{z\in[x,y]} \|x-y\|_{\nabla f^2(z)} \min_{z\in[x,y]} \|\nabla f(x) - \nabla f(y)\|_{[\nabla^2 f(z)]^{-1}}\,.\end{aligned}$$

Finally let $\alpha' = \arg\,\max_{\alpha\in[0,1]} \|x-y\|_{\nabla^2 f(z(\alpha))}$. We have

$$\begin{aligned}&\int_0^1 \|x-y\|^2_{\nabla^2 f(z(\alpha))}\, d\alpha = \int_0^{\alpha'} \|x-y\|^2_{\nabla^2 f(z(\alpha))}\, d\alpha + \int_{\alpha'}^1 \|x-y\|^2_{\nabla^2 f(z(\alpha))}\, d\alpha \\ &\geq \int_0^{\alpha'} \frac{\|x-y\|^2_{\nabla^2 f(z(\alpha'))}}{(1+(\alpha'-\alpha)M\|x-y\|_{\nabla^2 f(z(\alpha'))})^2}\, d\alpha + \int_{\alpha'}^1 \frac{\|x-y\|^2_{\nabla^2 f(z(\alpha'))}}{(1+(\alpha-\alpha')M\|x-y\|_{\nabla^2 f(z(\alpha'))})^2}\, d\alpha \\ &= \int_0^{\alpha'} \frac{\|x-y\|^2_{\nabla^2 f(z(\alpha'))}}{(1+\alpha M\|x-y\|_{\nabla^2 f(z(\alpha'))})^2}\, d\alpha + \int_0^{1-\alpha'} \frac{\|x-y\|^2_{\nabla^2 f(z(\alpha'))}}{(1+\alpha M\|x-y\|_{\nabla^2 f(z(\alpha'))})^2}\, d\alpha \\ &\geq \int_0^1 \frac{\|x-y\|^2_{\nabla^2 f(z(\alpha'))}}{(1+\alpha M\|x-y\|_{\nabla^2 f(z(\alpha'))})^2}\, d\alpha = \frac{\|x-y\|^2_{\nabla^2 f(z(\alpha'))}}{1+M\|x-y\|_{\nabla^2 f(z(\alpha'))}}\,,\end{aligned}$$

where the first inequality follows from

$$\max_{z\in[z(\alpha),z(\alpha'))} \left\|z(\alpha) - z(\alpha')\right\|_{\nabla^2 f(z)} = |\alpha-\alpha'| \max_{z\in[z(\alpha),z(\alpha'))} \|x-y\|_{\nabla^2 f(z)} = \|x-y\|_{\nabla^2 f(z(\alpha'))}$$

and applying Lemma 17.

Finally dividing both upper and lower bounds by

$$\max_{z\in[x,y]} \|x-y\|_{\nabla^2 f(z)} = \|x-y\|_{\nabla^2 f(z(\alpha'))}$$

completes the proof. ■

The following lemma follows directly from chaining Lemma 15 and Lemma 18.

Lemma 19 *For any M-self-concordant function f over $\mathcal{K} \subset \mathbb{R}^n$ and any $x, y \in \mathcal{K}$ and any $\xi < 1/2M$, it holds that*

$$\begin{aligned}&\min_{z\in[x,y]} \|\nabla f(x) - \nabla f(y)\|_{[\nabla^2 f(z)]^{-1}} = \xi \\ &\Rightarrow \frac{(1-2\xi M)^2}{(1-\xi M)^2}\nabla^2 f(y) \preceq \nabla^2 f(x) \preceq \frac{(1-\xi M)^2}{(1-2\xi M)^2}\nabla^2 f(y)\,.\end{aligned}$$

Appendix B. Quantum Learning: Preliminaries

B.1. Reductions for Online Learning of Quantum States with Log Loss

In this section we describe the online learning of quantum states problem as described in (Aaronson et al., 2018), and show that when the loss function is the log loss (and more generally, the KL-divergence), the problem can be cast in the form in Section 2. Recall that a *quantum state* on $\log_2 d$ qubits is a $d \times d$ Hermitian PSD matrix of trace 1. A *two-outcome measurement* is a $d \times d$ Hermitian matrix with eigenvalues in $[0, 1]$. When a quantum state X is measured using a two-outcome measurement E, the result is a Bernoulli random variable with probability of 1 being $\langle X, E\rangle$.

Aaronson et al. (2018) formulated the problem of online learning of quantum states as follows. In each round t, the learner constructs a quantum state X_t. In response, nature provides a two-outcome measurement E_t and a value $b_t \in [0, 1]$. The value b_t may be considered to be an approximation of $\langle X, E_t\rangle$ for some unknown quantum state X that we're trying to learn, or it can be thought of as the outcome in $\{0, 1\}$ of measuring the state X using E_t. However, as is standard in online learning, the pair (E_t, b_t) doesn't have to be consistent with any quantum state. The quality of the learner's prediction is given by a loss function $\ell : \mathbb{R}_+ \times [0, 1] \to \mathbb{R}$, and the loss in round t is computed as $\ell(\langle X_t, E_t\rangle, b_t)$. The goal is to minimize the regret, defined in the usual way as

$$\mathrm{Reg} = \sum_{t=1}^{T} \ell(\langle X_t, E_t\rangle, b_t) - \min_{\text{quantum state } X} \sum_{t=1}^{T} \ell(\langle X, E_t\rangle, b_t).$$

We now show that in either of the following two settings, the problem can be recast in the form given in Section 2.

Setting 1: $b_t \in \{0, 1\}$, and ℓ is the log loss, i.e. $\ell(p, b) = -\log(bp + (1 - b)(1 - p))$.
In this case, note that by setting the loss matrix to be $R_t := b_t E_t + (1 - b_t)(\mathbf{I}_d - E_t)$, we have $-\log(\langle X, R_t\rangle) = \ell(\langle X, E_t\rangle, b_t)$ for any quantum state X. This completes the reduction to the form in Section 2.

Setting 2: $b_t \in [0, 1]$, and ℓ is the KL-divergence, i.e. $\ell(p, b) = b \log(\frac{b}{p}) + (1 - b) \log(\frac{1-b}{1-p})$.
In each round t, sample a Bernoulli random variable y_t with probability of 1 being b_t. Then, setting $R_t = \frac{y_t}{b_t} E_t + \frac{1-y_t}{1-b_t}(\mathbf{I}_d - E_t)$, it is easy to check that $\mathbb{E}_{y_t} [-\log(\langle X, R_t\rangle)] = \ell(\langle X, E_t\rangle, b_t)$ for any quantum state X. This completes a *randomized* reduction to the form in Section 2. Note that setting 1 is the special case of this setting when $b_t \in \{0, 1\}$, and in this case the randomized reduction becomes deterministic and coincides with the reduction described for setting 1.

B.2. Preliminary Notation, Definitions and Useful Properties

In this section, we collect some basic notation, definitions and useful properties which allow for the extension of the usual concepts in online convex optimization to the case when the domain is Hermitian matrices.

Notation Recall that, we denote the set of $d \times d$ Hermitian matrices by $\mathcal{H}^d \subset \mathbb{C}^{d\times d}$, the set of $d \times d$ positive semi-definite Hermitian matrices by $\mathcal{H}^d_+$. Further, given two Hermitian matrices X, Y we define the standard inner product between them as $\langle X, Y\rangle := \mathrm{Tr}(X^*Y) = \mathrm{Tr}(XY)$. It is well-known, that the set of Hermitian matrices of fixed size constitute a finite dimensional real vector space (Olver et al., 2006). For any $M \in \mathcal{A}$, we denote $\overrightarrow{M} \in \mathbb{R}^{\tilde{d}}$ as the canonical representation such that $\langle X, Y \rangle = \langle \overrightarrow{X}, \overrightarrow{Y} \rangle$ and $\overrightarrow{\mathcal{A}} := \{\overrightarrow{M} \,|\, M \in \mathcal{A}\}$. In the quantum case, $\tilde{d} = \dim_{\mathcal{H}^d}$ and $\tilde{d} = d$ in the portfolio case.

In an overload of notation, any function $H : \mathcal{A} \to \mathbb{R}$ induces a function $H : \overrightarrow{\mathcal{A}} \to \mathbb{R}$ and we use $\overrightarrow{\nabla} H(X) \in \mathbb{R}^{\tilde{d}}$ to denote the gradient with respect to the function over this real valued vector space, while , $\nabla H(X) \in \mathcal{H}^d$ denotes the matrix representation of that gradient, and $\nabla^2 H(X) \in \mathbb{R}^{\tilde{d} \times \tilde{d}}$ is the Hessian in the real representation.

All functions in this paper are canonically defined over the cone $\mathcal{C} = \{c\overrightarrow{M} \,|\, M \in \mathcal{A}, c > 0\}$ induced by $\mathcal{A}$. By definition our action set satisfies $\overrightarrow{\mathcal{A}} = \{\overrightarrow{M} \in \mathcal{C} \,|\, \langle \overrightarrow{M}, \overrightarrow{\mathbf{I}_d} \rangle = 1\}$. For any twice differentiable function $H : \mathcal{C} \to \mathbb{R}$, we define the restriction to $\overrightarrow{\mathcal{A}}$ via the limit of the functions $H_\lambda(X) = H(X) + \lambda(\langle \overrightarrow{X}, \overrightarrow{\mathbf{I}_d} \rangle - 1)^2$, since

$$\lim_{\lambda \to \infty} H_\lambda(X) = \begin{cases} H(X) \text{ if } X \in \mathcal{A} \\ \infty \text{ otherwise.} \end{cases} .$$

For any $X, Y, Z \in \mathcal{A}$, we have $\nabla_\lambda H(Z) = \nabla H(Z)$. Further since $\nabla^2 H_\lambda(X) = \nabla^2 H(X) + \lambda \overrightarrow{\mathbf{I}_d}\overrightarrow{\mathbf{I}_d}^\top$, we have $\nabla^2_\lambda H(Z)(\overrightarrow{X} - \overrightarrow{Y}) = \nabla^2 H(Z)(\overrightarrow{X} - \overrightarrow{Y})$ and

$$\lim_{\lambda \to \infty} [\nabla^2 H_\lambda(X)]^{-1} = [\nabla^2 H(X)]^{-1} - \frac{[\nabla^2 H(X)]^{-1} \overrightarrow{\mathbf{I}_d}\overrightarrow{\mathbf{I}_d}^\top [\nabla^2 H(X)]^{-1}}{\overrightarrow{\mathbf{I}_d}^\top [\nabla^2 H(X)]^{-1} \overrightarrow{\mathbf{I}_d}} ,$$

which implies

$$\lim_{\lambda \to \infty} [\nabla^2 H_\lambda(X)]^{-1} \overrightarrow{\mathbf{I}_d} = 0 .$$

For any Hessian of functions $R, \hat{F}$ or G in the remainder of the upper bound sections, we refer to the restricted version obtained by $\lim_{\lambda \to \infty}$.

In the following lemma proves the Hessian of the $-\log\det()$ function is PD and lower bounded over the Hermitian subspace.

Lemma 20 *For any PSD matrix R, and PD matrix X, the matrix representation of the gradient of function $f(X) = -\log\det(X)$ is $\nabla f(X) := -X^{-1}$. Further the hessian satisfies the following properties*

- *For any PD X, and any Hermitian M, we have that*

$$\overrightarrow{M}^\top \nabla^2 f(X) \overrightarrow{M} = Tr(MX^{-1}MX^{-1}) .$$

- *For any PD X, and any Hermitian $M \neq 0$, we have that*

$$\overrightarrow{M}^* \nabla^2 f(X) \overrightarrow{M} > 0.$$

- *For any PD X, s.t. $Tr(X) \leq 1$, and any Hermitian M, we have that*

$$\overrightarrow{M}^* \nabla^2 f(X) \overrightarrow{M} \geq \|M\|^2.$$

- *For any PD X and any Hermitian M, we have that*

$$\overrightarrow{M}^\top[\nabla^2 f(X)]^{-1}\overrightarrow{M} = Tr(MXMX)\,.$$

- *f is 1 self-concordant.*

Proof We use the following results from Hjorungnes and Gesbert (2007) (Table 2), which show that the differential along the Hermitian matrices are equal to real symmetric ones. The differential of $\log\det(Z)$ is Tr$(Z^{-1}\,\boldsymbol{d}Z)$ and the differential of Z^{-1} is $-Z^{-1}\,\boldsymbol{d}ZZ^{-1}$. In our case, the differential is $\boldsymbol{d}Z = hM$ for a scalar h that goes to 0, and Z is evaluated at X, hence

$$\langle\overrightarrow{M}, \overrightarrow{\nabla} f(X)\rangle = \lim_{h\to 0} h^{-1}(-\log\det(X+hM) + \log\det(X)) = -\text{Tr}(MX^{-1})\,,$$

and

$$\overrightarrow{M}^\top\overrightarrow{\nabla}^2 f(X)\overrightarrow{M} = \lim_{h\to 0} h^{-1}(-\text{Tr}(M(X+hM)^{-1}) + \text{Tr}(MX^{-1})) = \text{Tr}(MX^{-1}MX^{-1})\,.$$

Next we have that

$$\overrightarrow{M}^\top\nabla^2 f(X)\overrightarrow{M} = \text{Tr}(MX^{-1}MX^{-1}) = \text{Tr}((X^{-1/2}MX^{-1/2})^2) = \sum_i \lambda_i^2(X^{-1/2}MX^{-1/2}) > 0,$$

where the last inequality follows because we know that M is Hermitian and not identically 0. Next we show that for any Hermitian matrix M and any PD matrix X such that Tr$(X) \leq 1$, we have that

$$\overrightarrow{M}^\top\nabla^2 f(X)\overrightarrow{M} \geq \|\overrightarrow{M}\|^2.$$

This is proved as follows. By the spectral theorem, we can write X^{-1} as $U^*\operatorname{diag}(\Lambda)U$, where U is a unitary matrix, and Λ is the vector of eigenvalues of X^{-1}, which are all at least 1 since Tr$(X) \leq 1$. Now we have

$$\begin{aligned}\overrightarrow{M}^\top\nabla^2 f(X)\overrightarrow{M} = \text{Tr}(MX^{-1}MX^{-1}) &= \text{Tr}(MU^*\operatorname{diag}(\Lambda)UMU^*\operatorname{diag}(\Lambda)U)\\ &= \text{Tr}(\tilde{M}\operatorname{diag}(\Lambda)\tilde{M}\operatorname{diag}(\Lambda)),\end{aligned}$$

where $\tilde{M} := UMU^*$. Now consider the function $f : \mathbb{R}_+^d \to \mathbb{R}$ defined as

$$f(\lambda) = \text{Tr}(\tilde{M}\operatorname{diag}(\lambda)\tilde{M}\operatorname{diag}(\lambda)).$$

An easy calculation using the fact that $\tilde{M}$ is Hermitian yields, for any $i \in [d]$,

$$\frac{\partial f(\lambda)}{\partial\lambda_i} = \sum_{k\neq i}|\tilde{M}_{ik}|^2\lambda_k + 2|\tilde{M}_{ii}|^2\lambda_i \geq 0.$$

Since $\Lambda \geq \mathbf{1}$ entrywise, the above inequality implies that

$$\overrightarrow{M}^*\nabla^2 f(X)\overrightarrow{M} = f(\Lambda) \geq f(\mathbf{1}) = \text{Tr}(\tilde{M}^2) = \|M\|^2.$$

Next, the inverse Hessian is well defined since the function has just been shown to be strongly convex. The convex conjugate f^* is defined over $-\mathcal{H}$ and is given by

$$f^*(Y) = \sup_{X\in\mathcal{H}} \langle X, Y\rangle - f(X) = -d + \log\det(-Y^{-1}) = -d - f(-Y)\,,$$

as can be seen by setting the derivative to 0 in the optimization over X. Since $\nabla^2 f^*(\nabla f(X)) = [\nabla^2 f(X)]^{-1}$(Rockafellar, 2015), the statement follows from the first property. Finally for self-concordancy, we have just shown

$$\nabla^2 f(X)[\overrightarrow{M}, \overrightarrow{M}] = \text{Tr}(MX^{-1}MX^{-1}) = \text{Tr}((X^{-\frac{1}{2}}MX^{-\frac{1}{2}})^2)\,,$$

from which follows

$$\begin{aligned}\nabla^3 f(X)[\overrightarrow{M}, \overrightarrow{M}, \overrightarrow{M}] &= \lim_{h\to 0} h^{-1}(\text{Tr}(M(X+hM)^{-1}M(X+hM)^{-1}) - \text{Tr}(MX^{-1}MX^{-1}))\\ &= 2\text{Tr}(MX^{-1}MX^{-1}MX^{-1})\\ &= 2\text{Tr}((X^{-\frac{1}{2}}MX^{-\frac{1}{2}})^3)\,.\end{aligned}$$

Denote $(\lambda_i)_{i=1}^d$ the eigenvalues of the matrix $X^{-\frac{1}{2}}MX^{-\frac{1}{2}}$, then we have

$$|\nabla^3 f(X)[\overrightarrow{M}, \overrightarrow{M}, \overrightarrow{M}]| = 2|\sum_{i=1}^d \lambda_i^3| \le 2\sum_{i=1}^d |\lambda_i|^3 \le 2(\sum_{i=1}^d |\lambda_i|^2)^{\frac{3}{2}} = 2\left\|\overrightarrow{M}\right\|^3_{\nabla^2 f(X)}\,.$$

■

Lemma 21 *Any function of the form $H(X) = \eta^{-1}R(X) + Q(X)$ for a quadratic Q with PSD hessian is $\sqrt{\eta}$ self-concordant.*

Proof The quadratic part only affects the RHS in the definition of self-concordancy, hence it is sufficient to show the statement for $\eta^{-1}R(X)$. This follows directly from the last point in Lemma 20 and 16. ■

The definition of surrogate functions and the biased surrogate functions are naturally extended to the quantum learning case from the definitions provided in (2), (3). We now recall the definitions provided in the algorithm description,

$$G_\tau^e(X) := \sum_{s=1}^{\tau} g_s^e(X) + \eta^{-1}R(X) \quad \text{and} \quad {X_{\tau+1}^\star}^e := \arg\min_{X\in\mathcal{A}} G_\tau^e(X).$$

$$\hat{F}_\tau^e(X) := \sum_{s=1}^{\tau} \hat{f}_s^e(X) + \eta^{-1}R(X) \quad \text{and} \quad {U_{\tau+1}^\star}^e := \arg\min_{X\in\mathcal{A}} \hat{F}_\tau^e(X).$$

The following lemma establishes some conditions on the minimizers. We drop the notation for epoch e from these statements for brevity.

Lemma 22 *We have that the following statements hold for all τ,*

- $X^\star_{\tau+1} \succ 0, U^\star_{\tau+1} \succ 0$, *i.e. lie in the interior of the action set* $\mathcal{A}$
- *Given any Hermitian matrices* X, U *such that* $Tr(X) = Tr(U) = 0$, *we have that*

$$\langle \nabla G_\tau(X^\star_{\tau+1}), X\rangle = 0 \text{ and } \langle \nabla \hat{F}_\tau(U^\star_{\tau+1}), U\rangle = 0.$$

- *Further we have that*

$$[\overrightarrow{\nabla}^2 \hat{F}_\tau(U^\star_{\tau+1})]^{-1} \overrightarrow{\nabla} \hat{F}_\tau(U^\star_{\tau+1}) = 0$$
$$[\overrightarrow{\nabla}^2 \hat{G}_\tau(X^\star_{\tau+1})]^{-1} \overrightarrow{\nabla} \hat{G}_\tau(X^\star_{\tau+1}) = 0$$

Proof The first statement is immediate by noting that for any $X \succeq 0$ with at least one eigenvalue approaching 0, we have that $R(X)$ and thus G_τ approaches ∞ and for all PD matrices $\in \mathcal{A}$, G_τ is finite.

For the second statement we will prove the first inequality. The proof for the second inequality is analogous. We assume $X \neq 0$, otherwise the statement is immediate. Since $X^\star_{\tau+1} \succ 0$, there exists a $\delta > 0$ such that $X^\pm = X^\star_{\tau+1} \pm \delta X$ such that $X^+, X^- \in \mathcal{A}$. Consider the function $X(\alpha) := X^\star_{\tau+1} + \alpha\delta X$ over $\alpha \in [-1, 1]$. The function $G_\tau(X(\alpha))$ is continuously differentiable in α and the derivative at $\alpha = 0$ is via chain rule

$$\left|\frac{\partial}{\partial\alpha} G_\tau(X(\alpha))\right|_{\alpha=0} = \langle \nabla G_\tau(X^\star_{\tau+1}), X\rangle .$$

Since $X(0) = X^\star_{\tau+1}$ is the minimizer, this must be 0. The last statement follows immediately from the previous calculation and that $\overrightarrow{\mathbf{I}_d}$ is in the kernel of the Hessian inverse. ■

We provide the proof of the following lemma whose restriction over the reals is well-known and is used repeatedly in the proofs of Online Newton Step like algorithms.

Lemma 23 *Given a sequence of PD matrices* $X_1 \preceq X_2 \preceq X_3 \ldots X_T$, *we have that*

$$\sum_{t=1}^{T-1} \langle X^{-1}_{t+1}, X_{t+1} - X_t\rangle \leq \log(\det(X_T)) - \log(\det(X_1)).$$

Proof We first begin by providing the proof of a simpler statement which implies the above statement via a simple summation. Given two PD matrices $X \preceq Y$, we have that

$$\langle Y^{-1}, Y - X\rangle \leq \log(\det(Y)) - \log(\det(X)).$$

To prove the above we consider the following function $\phi(\alpha)$ defined as

$$\phi(\alpha) := \log\det(\alpha Y + (1-\alpha)X).$$

Using Lemma 20 we see that ϕ is a concave function over α and that

$$\phi'(\alpha) = \langle Y - X, (\alpha Y + (1-\alpha)X)^{-1}\rangle.$$

Therefore using concavity we have that $\phi'(1) \leq \phi(1) - \phi(0)$ which implies the requisite statement by substitution.

■

Appendix C. Algorithm for Quantum Learning with Log Loss

Algorithm 3: SCHRÖDINGER'S-BISONS

input: T, B, η, β.

initialize: $\forall e \in \mathbb{N}: \; P_0^e = d\mathbf{I}_d, G_0^e(\cdot) = \hat{F}_0^e(\cdot) = \eta^{-1}R(\cdot), X_1^e = U_1^e = \arg\min_{X \in \mathcal{A}} G_0^e(X)$.

$e \leftarrow 1, \tau \leftarrow 1$

for $t = 1, \ldots$ **do**

  $f_t \leftarrow$ receive from playing $X_t \leftarrow X_\tau^e$.

  $\hat{f}_\tau^e = \hat{f}_t \leftarrow$ construct according to (2).

  $\hat{F}_\tau^e \leftarrow \hat{F}_{\tau-1}^e + \hat{f}_\tau^e$

  $G_\tau^e \leftarrow G_{\tau-1}^e + g_\tau^e$, where $g_\tau^e(X) := \hat{f}_\tau^e(X) - \langle X, P_\tau^e - P_{\tau-1}^e \rangle B$.

  $X_{\tau+1}^e \leftarrow \text{APPROX-SOLVE}_X(G_\tau^e(\cdot), X_\tau^e), U_{\tau+1}^e \leftarrow \text{APPROX-SOLVE}_U(\hat{F}_\tau^e(\cdot), U_\tau^e)$

  $P_{\tau+1}^e = P_\tau^e + [X_{\tau+1}^e]^{-\frac{1}{2}} \left(\mathbf{I}_d - [X_{\tau+1}^e]^{\frac{1}{2}} P_\tau^e [X_{\tau+1}^e]^{\frac{1}{2}} \right)_+ [X_{\tau+1}^e]^{-\frac{1}{2}}$

  if $U_{\tau+1}^e \not\succcurlyeq \frac{1}{2(1+6\eta)\beta}[P_{\tau+1}^e]^{-1}$ **then**

    $e \leftarrow e+1, \tau \leftarrow 1$ `// Reset the algorithm`

  else

    $\tau \leftarrow \tau + 1$

  end

end

Appendix D. Preliminary definitions and properties

In this section we provide some general definitions and other properties necessary for the analysis of the BISONS algorithm. Given a PSD matrix $A \in \mathcal{H}_+^d$, we associate a norm over $\mathcal{H}_+^d$, defined for any $W \in \mathcal{H}_+^d$ as

$$\|W\|_A = \sqrt{\text{Tr}(WAWA)} = \sqrt{\text{Tr}((A^{1/2}WA^{1/2})^2)}\,.$$

Lemma 24 *For any positive semi-definite A, $\|\cdot\|_A$ is a pseudo-norm.*

Proof The only non-trivial property is the triangle inequality. We have

$$\begin{aligned}\|V+W\|_A^2 &= \text{Tr}((A^{\frac{1}{2}}(V+W)A^{\frac{1}{2}})^2) = \|V\|_A^2 + \|W\|_A^2 + 2\text{Tr}((A^{\frac{1}{2}}VA^{\frac{1}{2}})(A^{\frac{1}{2}}WA^{\frac{1}{2}})) \\ &\le \|V\|_A^2 + \|W\|_A^2 + 2\sqrt{\text{Tr}((A^{\frac{1}{2}}VA^{\frac{1}{2}})^2)\text{Tr}((A^{\frac{1}{2}}WA^{\frac{1}{2}})^2)} = (\|V\|_A + \|W\|_A)^2\,,\end{aligned}$$

where the inequality is due to the fact that for PSD matrices A, B, we have that $\text{Tr}(AB) \le \sqrt{\text{Tr}(A^2)\text{Tr}(B^2)}$, which follows from the Cauchy-Schwarz inequality. ∎

Lemma 25 *For any PD matrices A, B such that $\|A-B\|_{B^{-1}} \le \lambda$ for some $\lambda \ge 0$, the eigenvalues of $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ lie within the interval $[1-\lambda, 1+\lambda]$.*

Proof We have

$$\begin{aligned}\|A-B\|_{B^{-1}}^2 &= \mathrm{Tr}((A-B)B^{-1}(A-B)B^{-1})\\ &= \mathrm{Tr}((B^{-\frac{1}{2}}AB^{-\frac{1}{2}}-I)^2)\\ &= \sum_{i=1}^d \left(\mathrm{ev}_i(B^{-\frac{1}{2}}AB^{-\frac{1}{2}})-1\right)^2 \leq \lambda^2,\end{aligned}$$

where ev_i represents the i^{th} eigenvalue. Therefore every eigenvalue satisfies

$$|\mathrm{ev}_i(B^{-\frac{1}{2}}AB^{-\frac{1}{2}})-1| \leq \lambda\,.$$

■

For any PSD matrix A, B define

$$[A,B] := \{\alpha A + (1-\alpha)B \,|\, \alpha \in [0,1]\}. \tag{8}$$

Lemma 26 *For any PD matrix A, B such that*

$$\max_{C\in[A,B]} \|A-B\|_{C^{-1}} \leq \lambda\,,$$

for some $\lambda \geq 0$. Then it holds for all $D, E \in [A,B]$:

$$D \preceq (1+\lambda)E\,, \quad D^{-1} \preceq (1+\lambda)E^{-1}\,.$$

Proof Since $D, E \in [A,B]$, there exists $c:\ |c| \leq 1$ such that $D-E = c(A-B)$. Hence

$$\|D-E\|_{D^{-1}} \leq \lambda\,.$$

Applying Lemma 25 completes the first part. Repeating the same argument, but now starting with $\|D-E\|_{E^{-1}} \leq \lambda$, yields the second claim. ■

Lemma 27 *For any function $H(\cdot) = \eta^{-1}R(\cdot)+Q(\cdot)$ for a quadratic function Q with PSD Hessian, it holds that for any $A, B \in \mathcal{A}$ and for any $\xi \geq 0$ and for any $D, E \in [A,B]$:*

$$\max_{C\in[A,B]} \left\|\overrightarrow{A}-\overrightarrow{B}\right\|_{\overrightarrow{\nabla}^2 H(C)} \leq \xi \Rightarrow (1+\sqrt{\eta}\xi)^{-1}D \preceq E \preceq (1+\sqrt{\eta}\xi)D\,.$$

Proof We have

$$\max_{C\in[A,B]} \|A-B\|_{C^{-1}} = \sqrt{\eta}\max_{C\in[A,B]} \left\|\overrightarrow{A}-\overrightarrow{B}\right\|_{\eta^{-1}\overrightarrow{\nabla}^2 R(C)} \leq \sqrt{\eta}\max_{C\in[A,B]} \left\|\overrightarrow{A}-\overrightarrow{B}\right\|_{\overrightarrow{\nabla}^2 H(C)} \leq \sqrt{\eta}\xi\,.$$

Applying Lemma 26 completes the proof. ■

Appendix E. BISONS detailed analysis

In this section we provide the details for the analysis of our algorithms 1, 3, eventually proving Theorems 1 and 5. Before delving into the analysis we request the reader to familiarize themselves with the requisite notation, definition and properties listed out in Sections D, C. Since BISONS is a special case of SCHRÖDINGER'S-BISONS, we will provide the analysis focused on the quantum learning case, i.e. the domain will be PSD Hermitian matrices, however all the statements will hold when these matrices are real and diagonal as will be the case for the online optimal portfolio.

We first provide a proof of Lemma 7. We further begin the core analysis by providing some useful auxiliary lemmas and the lemmas governing the stability of the output of the algorithm in the next two subsections. We will restrict attention in the next two subsections to any fixed epoch and there for brevity we will remove the epoch superscript e, from the lemma statements as well as proofs. All the statements should be understood to hold for any particular epoch.

E.1. Proof of Lemma 7

Proof Equality at $x = y_t$ holds by construction. We have $h'(x) = -x^{-1}$, which is concave and $\underline{\hat{h}'_t}(x) = \min\{-(1+\beta)y_t^{-1} + \beta x y_t^{-2}, -\beta y_t^{-1}\}$, which is piece-wise linear. A quick calculation shows $h'(y_t) = \underline{h'_t}(y_t)$ and $h'(\beta^{-1}y_t) = \underline{h'_t}(\beta^{-1}y_t)$. Hence for $x < y_t$, we have $h'(x) < \underline{h'_t}(x)$ and for $\beta^{-1}y_t \geq x > y_t$, we have $h'(x) > \underline{h'_t}(x)$. Finally, the derivative of $h'(x)$ is monotonically increasing which implies $h'(x) > \underline{h'_t}(x)$ for $x > \beta^{-1}y_t$, which completes the proof. ■

E.2. Auxiliary Lemmas

In this section we collect some basic lemmas regarding the matrices X_τ, P_τ generated by the algorithm. We recall the definition of P_τ defined in (7) as

$$P_\tau := P_{\tau-1} + X_\tau^{-\frac{1}{2}} \left(\mathbf{I}_d - X_\tau^{\frac{1}{2}} P_{\tau-1} X_\tau^{\frac{1}{2}}\right)_+ X_\tau^{-\frac{1}{2}},$$

which in particular implies that for all τ,

$$X_{\tau+1}^{\frac{1}{2}}(P_{\tau+1} - P_\tau)X_{\tau+1}^{\frac{1}{2}} = \left(\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}}\right)_+ .$$

The next two lemmas state the main properties satisfied by our choice of P_τ. These properties prompt the choice of the definition for P_τ.

Lemma 28 *We have that for all τ,*

$$P_\tau \succeq P_{\tau-1} \quad \textit{and} \quad P_\tau \succeq X_\tau^{-1}.$$

Proof The first statement is immediate from the definition of P_τ. For the second inequality note that

$$X_\tau^{\frac{1}{2}} P_\tau X_\tau^{\frac{1}{2}} = X_\tau^{\frac{1}{2}} P_{\tau-1} X_\tau^{\frac{1}{2}} + \left(\mathbf{I}_d - X_\tau^{\frac{1}{2}} P_{\tau-1} X_\tau^{\frac{1}{2}}\right)_+ \succeq \mathbf{I}_d, \tag{9}$$

which implies that $P_\tau \succeq X_\tau^{-1}$. ■

Lemma 29 *For any τ, we have*

$$\langle X_{\tau+1}, P_{\tau+1} - P_\tau \rangle = \langle P_{\tau+1}^{-1}, P_{\tau+1} - P_\tau \rangle.$$

Proof Recall, by definition

$$P_{\tau+1} = P_\tau + X_{\tau+1}^{-\frac{1}{2}} \left(\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}} \right)_+ X_{\tau+1}^{-\frac{1}{2}} .$$

Hence

$$\langle X_{\tau+1}, P_{\tau+1} - P_\tau \rangle = \mathrm{Tr}\left(\left(\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}} \right)_+ \right) = \sum_{i=1}^d \max\{1 - \lambda_i, 0\} ,$$

where λ_i are the eigenvalues of $X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}}$. For the RHS, we have

$$\langle P_{\tau+1}^{-1}, P_{\tau+1} - P_\tau \rangle = \langle (X_{\tau+1}^{\frac{1}{2}} P_{\tau+1} X_{\tau+1}^{\frac{1}{2}})^{-1}, X_{\tau+1}^{\frac{1}{2}} (P_{\tau+1} - P_\tau) X_{\tau+1}^{\frac{1}{2}} \rangle .$$

Note that $(X_{\tau+1}^{\frac{1}{2}} P_{\tau+1} X_{\tau+1}^{\frac{1}{2}}) = X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}} + \left(\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}} \right)_+$ modifies the eigenvalues of $X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}}$ such that they are lower bounded by 1. Therefore

$$\langle (X_{\tau+1}^{\frac{1}{2}} P_{\tau+1} X_{\tau+1}^{\frac{1}{2}})^{-1}, X_{\tau+1}^{\frac{1}{2}} (P_{\tau+1} - P_\tau) X_{\tau+1}^{\frac{1}{2}} \rangle = \sum_{i=1}^d \frac{\max\{1 - \lambda_i, 0\}}{\max\{1, \lambda_i\}} = \sum_{i=1}^d \max\{1 - \lambda_i, 0\} ,$$

where the last equality follows from the nominator being non-zero only if the denominator is 1. ■

The following is a useful lemma we collect here.

Lemma 30 *For any τ, it holds that*

$$\|P_{\tau+1} - P_\tau\|_{X_{\tau+1}} \leq \|X_{\tau+1} - X_\tau\|_{X_\tau^{-1}} .$$

Proof Denote $\tilde{D} = X_{\tau+1}^{\frac{1}{2}} (P_{\tau+1} - P_\tau) X_{\tau+1}^{\frac{1}{2}}$, therefore we have that

$$\tilde{D} = \left(\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}} \right)_+ .$$

We have

$$\begin{aligned}
\|P_{\tau+1} - P_\tau\|_{X_{\tau+1}}^2 = \mathrm{Tr}(\tilde{D}^2) &= \mathrm{Tr}\left((\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}})_+^2 \right) \\
&\leq \mathrm{Tr}\left((\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} X_\tau^{-1} X_{\tau+1}^{\frac{1}{2}})_+^2 \right) \\
&\leq \mathrm{Tr}\left((\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}} X_\tau^{-1} X_{\tau+1}^{\frac{1}{2}})^2 \right) \\
&= \|X_{\tau+1} - X_\tau\|_{X_\tau^{-1}}^2 ,
\end{aligned}$$

where the first inequality uses the fact $P_\tau \succeq X_\tau^{-1}$ from Lemma 28. ■

Finally as a result of our reset condition we have the following lemma.

Lemma 31 *Let $s \le \tau$ be two time indices belonging to the same epoch, such that the reset condition was not triggered up to time index $\tau - 1$. Then we have that*

$$U_\tau^\star \preceq 2U_\tau \preceq \beta^{-1} X_s\,.$$

Proof Due to the reset condition, we know that $U_\tau \preceq (2(1+6\eta)\beta)^{-1} P_\tau^{-1} \preceq (2\beta)^{-1} P_\tau^{-1}$. Further since by Lemma 28, $P_\tau^{-1} \preceq X_s$, the second inequality of the lemma follows. By requirement of the approximated solution, we have

$$\left\| \overrightarrow{\nabla} \hat{F}_{\tau-1}(U_\tau) \right\|_{(\overrightarrow{\nabla}^2 \hat{F}_{\tau-1}(U_\tau))^{-1}} \le \frac{1}{2\sqrt{\eta}}\,.$$

Note by Lemma 21, the function F_τ is $\sqrt{\eta}$-self concordant and therefore by Lemma 18, this implies that

$$\|U_\tau - U_\tau^\star\|_{(\overrightarrow{\nabla}^2 \hat{F}_{\tau-1}(U_\tau))^{-1}} \le \frac{1}{\sqrt{\eta}}.$$

Further using Lemma 27 and Lemma 25 yields $U_\tau^\star \preceq 2U_\tau$. ■

E.3. Stability Lemmas

In this section we show that successive iterates X_τ, U_τ, P_τ and $X_{\tau+1}, U_{\tau+1}, P_{\tau+1}$ do not move too far away from each other due to the log barrier, establishing the requisite stability of our method. These results are summarized in lemmas 33 and 34. Our stability lemmas hold under the following constraints over the algorithm parameters η, β.

$$\eta \le \min\{\frac{1}{4B}, \frac{\beta}{4}, \frac{1}{63}\} \tag{10}$$

$$\beta \le 2/\sqrt{3} - 1 \tag{11}$$

$$T \ge \max\{2d, \beta^{-1}\} \tag{12}$$

As a reminder by Assumption 1 we have that following properties,

$$\left\| \overrightarrow{\nabla} G_\tau(X_{\tau+1}) \right\|_{[\overrightarrow{\nabla}^2 G_\tau(X_{\tau+1})]^{-1}} \le \min\{6\eta, 6\sqrt{\eta} \left\| \overrightarrow{\nabla} G_\tau(X_\tau) \right\|_{[\overrightarrow{\nabla}^2 G_\tau(X_\tau)]^{-1}}\}$$

$$\left\| \overrightarrow{\nabla} \hat{F}_\tau(U_{\tau+1}) \right\|_{[\overrightarrow{\nabla}^2 \hat{F}_\tau(U_{\tau+1})]^{-1}} \le \frac{1}{8\sqrt{\eta}}$$

Lemma 32 *For any τ, it holds that*

$$X_{\tau+1}^{-1} \preceq (1+\lambda) X_\tau^{-1} \quad \Rightarrow \quad P_{\tau+1} \preceq (1+\lambda) P_\tau\,.$$

Proof We assume LHS above is true. By Lemma 28 we have that,

$$X_{\tau+1}^{-1} \preceq (1+\lambda) X_\tau^{-1} \preceq (1+\lambda) P_\tau\,.$$

Therefore we have that,

$$\mathbf{I}_d \preceq (1+\lambda) X_{\tau+1}^{\frac{1}{2}} P_\tau X_{\tau+1}^{\frac{1}{2}}\,.$$

Hence by definition of $P_{\tau+1}$,

$$X_{\tau+1}^{\frac{1}{2}}(P_{\tau+1}-(1+\lambda)P_\tau)X_{\tau+1}^{\frac{1}{2}} = (\mathbf{I}_d - X_{\tau+1}^{\frac{1}{2}}P_\tau X_{\tau+1}^{\frac{1}{2}})_+ - \lambda X_{\tau+1}^{\frac{1}{2}}P_\tau X_{\tau+1}^{\frac{1}{2}}$$
$$\preceq (\mathbf{I}_d - (1+\lambda)X_{\tau+1}^{\frac{1}{2}}P_\tau X_{\tau+1}^{\frac{1}{2}})_+ = 0\,.$$

Finally, this implies

$$(P_{\tau+1}-(1+\lambda)P_\tau) \preceq 0 \quad \Leftrightarrow \quad P_{\tau+1} \preceq (1+\lambda)P_\tau$$

as claimed. ■

Lemma 33 *If η satisfies constraint (10), then for any $t \in [T]$:*

$$\frac{1}{1+6\eta}X_{\tau+1} \preceq X_\tau \preceq (1+6\eta)X_{\tau+1}\,,$$

as well as

$$P_{\tau+1} \preceq (1+6\eta)P_\tau\,.$$

Proof We prove the statement for X, for which the claim for P follows directly via Lemma 32. The proof follows by induction. Set by convention $X_0 = X_1 = \frac{1}{d}\mathbf{I}_d$, then the condition holds for $\tau = 0$. Further we will bound the following quantity quantity,

$$\min_{\Xi\in[X_\tau,X_{\tau+1}]}\left\|\overrightarrow{\nabla}G_\tau(X_\tau)-\overrightarrow{\nabla}G_\tau(X_{\tau+1})\right\|_{[\overrightarrow{\nabla}^2G_\tau(\Xi)]^{-1}} \le \left\|\overrightarrow{\nabla}G_\tau(X_\tau)\right\|_{[\overrightarrow{\nabla}^2G_\tau(X_\tau)]^{-1}} + \left\|\overrightarrow{\nabla}G_\tau(X_{\tau+1})\right\|_{[\overrightarrow{\nabla}^2G_\tau(X_\tau)]^{-1}}\,.$$

The first term is bounded by

$$\begin{aligned}
\left\|\overrightarrow{\nabla}G_\tau(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}} &\le \left\|\overrightarrow{\nabla}G_{\tau-1}(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}} + \left\|\overrightarrow{\nabla}g_\tau(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}}\\
&\le \left\|\overrightarrow{\nabla}G_{\tau-1}(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_{\tau-1}(X_\tau))^{-1}} + \left\|\overrightarrow{\nabla}g_\tau(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}}\\
&\le 6\eta + \left\|\overrightarrow{\nabla}\hat{f}_\tau(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}} + B\left\|\overrightarrow{P}_\tau-\overrightarrow{P}_{\tau-1}\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}}\\
&\le 6\eta + \left\|\overrightarrow{\nabla}\hat{f}_\tau(X_\tau)\right\|_{\eta(\overrightarrow{\nabla}^2R(X_\tau))^{-1}} + B\left\|\overrightarrow{P}_\tau-\overrightarrow{P}_{\tau-1}\right\|_{\eta(\overrightarrow{\nabla}^2R(X_\tau))^{-1}}\\
&\le 6\eta + \sqrt{\eta}\left(\left\|\nabla\hat{f}_\tau(X_\tau)\right\|_{X_\tau} + B\left\|P_\tau-P_{\tau-1}\right\|_{X_\tau}\right)\\
&\le 6\eta + \sqrt{\eta}\left(\frac{\sqrt{\mathrm{Tr}((X_\tau^{\frac{1}{2}}R_\tau X_\tau^{\frac{1}{2}})^2)}}{\mathrm{Tr}(X_\tau^{\frac{1}{2}}R_\tau X_\tau^{\frac{1}{2}})} + B\left\|X_\tau - X_{\tau-1}\right\|_{X_\tau^{-1}}\right)\\
&\le 6\eta + \sqrt{\eta}(1+6\eta B) \le 3\sqrt{\eta}\,.
\end{aligned} \tag{13}$$

The third inequality above follows via the approximation assumption on the approximate solver, the second last equation via the definition of the relevant quantities and the inequality follows from the constraints on η defined in (10) and induction assumption. Note from the optimality of $X_{\tau+1}^\star$ that

$$\left\|\overrightarrow{\nabla}G_\tau(X_\tau)-\overrightarrow{\nabla}G_\tau(X_{\tau+1}^\star)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}} = \left\|\overrightarrow{\nabla}G_\tau(X_\tau)\right\|_{(\overrightarrow{\nabla}^2G_\tau(X_\tau))^{-1}}\,. \tag{14}$$

Combining the derivations above with Lemma 19 we get that

$$\overrightarrow{\nabla}^2 G_\tau(X_\tau) \preceq \frac{1}{(1-6\eta)^2}\overrightarrow{\nabla}^2 G_\tau(X^\star_{\tau+1})\,. \tag{15}$$

By the condition on the approximate solver, we have

$$\|\nabla G_\tau(X_{\tau+1})\|_{(\overrightarrow{\nabla}^2 G_\tau(X_{\tau+1}))^{-1}} \le 6\eta\,,$$

hence using (14) and Lemma 19 we have that,

$$\overrightarrow{\nabla}^2 G_\tau(X^\star_{\tau+1}) \preceq \frac{1}{(1-12\eta^{\frac{3}{2}})^2}\overrightarrow{\nabla}^2 G_\tau(X_{\tau+1})\,.$$

Combining everything yields

$$\min_{\Xi\in[X_\tau,X_{\tau+1}]}\left\|\overrightarrow{\nabla} G_\tau(X_\tau)-\overrightarrow{\nabla} G_\tau(X_{\tau+1})\right\|_{[\overrightarrow{\nabla}^2 G_\tau(\Xi)]^{-1}} \le 3\sqrt{\eta}+\frac{6\eta}{(1-6\eta)(1-12\eta^{\frac{3}{2}})} \le 4\sqrt{\eta}\,.$$

Note by Lemma 21, the function G_τ is $\sqrt{\eta}$-self concordant and therefore chaining Lemma 18 and 27 yields

$$\lambda^{-1}X_{\tau+1} \preceq X_\tau \preceq \lambda X_{\tau+1}$$

for

$$\lambda = 1+\frac{4\eta}{1-4\eta} \le 1+6\eta\,.$$

■

Lemma 34 *If η and β satisfy constraints (10) and (11) and no reset is triggered at time $\tau-1$, then*

$$U_{\tau+1}\preceq 2U_\tau \quad \textit{and} \quad U^\star_{\tau+1}\preceq 2U_\tau.$$

Proof We begin by bounding the respective norm

$$\min_{\Xi\in[U_\tau,U_{\tau+1}]}\left\|\overrightarrow{\nabla} F_\tau(U_\tau)-\overrightarrow{\nabla} F_\tau(U_{\tau+1})\right\|_{[\overrightarrow{\nabla}^2 F_\tau(\Xi)]^{-1}} \le \left\|\overrightarrow{\nabla} F_\tau(U_\tau)\right\|_{[\overrightarrow{\nabla}^2 F_\tau(U_\tau)]^{-1}} + \left\|\overrightarrow{\nabla} F_\tau(U_{\tau+1})\right\|_{[\overrightarrow{\nabla}^2 F_\tau(U_\tau)]^{-1}}\,.$$

The first term is

$$\begin{aligned}\left\|\overrightarrow{\nabla} F_\tau(U_\tau)\right\|_{(\overrightarrow{\nabla}^2 F_\tau(U_\tau))^{-1}} &\le \left\|\overrightarrow{\nabla} F_{\tau-1}(U_\tau)\right\|_{(\overrightarrow{\nabla}^2 F_\tau(U_\tau))^{-1}} + \left\|\overrightarrow{\nabla} \hat{f}_\tau(U_\tau)\right\|_{(\overrightarrow{\nabla}^2 F_\tau(U_\tau))^{-1}}\\ &\le \frac{1}{8\sqrt{\eta}}+\sqrt{\eta}\left\|\nabla \hat{f}_\tau(U_\tau)\right\|_{U_\tau} && \text{(Lemma 20)}\end{aligned}$$

Recall $\nabla \hat{f}_\tau(U_\tau) = -\left(1+\beta-\beta\frac{\langle U_\tau,R_\tau\rangle}{\langle X_\tau,R_\tau\rangle}\right)\frac{R_\tau}{\langle X_\tau,R_\tau\rangle}$. Since no reset is triggered at time $\tau-1$, we have using Lemma 31 that $0\le\frac{\langle U_\tau,R_\tau\rangle}{\langle X_\tau,R_\tau\rangle}\le\frac{1}{2\beta}$. Therefore we have that

$$\left\|\nabla \hat{f}_\tau(U_\tau)\right\|_{U_\tau} = \left(1+\beta-\beta\frac{\langle U_\tau,R_\tau\rangle}{\langle X_\tau,R_\tau\rangle}\right)\frac{\sqrt{\mathrm{Tr}(U_\tau R_\tau U_\tau R_\tau)}}{\langle X_\tau,R_\tau\rangle} \le \left(1+\beta-\beta\frac{\langle U_\tau,R_\tau\rangle}{\langle X_\tau,R_\tau\rangle}\right)\frac{\langle U_\tau,R_\tau\rangle}{\langle X_\tau,R_\tau\rangle}.$$

Maximizing the above expression over all choice of $\frac{\langle U_\tau, R_\tau\rangle}{\langle X_\tau, R_\tau\rangle} \in [0, 1/2\beta]$ we get that

$$\left\|\nabla \hat{f}_\tau(U_\tau)\right\|_{U_\tau} \leq \frac{(1+\beta)^2}{4\beta} \leq \frac{1}{3\beta},$$

which follows by the constraint on β in (11). This results in

$$\left\|\overrightarrow{\nabla} F_\tau(U_\tau)\right\|_{(\overrightarrow{\nabla}^2 F_\tau(U_\tau))^{-1}} \leq \frac{1}{\sqrt{\eta}}\left(\frac{1}{8} + \frac{\eta}{3\beta}\right) \leq \frac{5}{24\sqrt{\eta}}\,. \tag{16}$$

Further using the optimiality of $U^\star_{\tau+1}$ we have that

$$\left\|\overrightarrow{\nabla} F_\tau(U_\tau) - \overrightarrow{\nabla} F_\tau(U^\star_{\tau+1})\right\|_{(\overrightarrow{\nabla}^2 F_\tau(U_\tau))^{-1}} = \left\|\overrightarrow{\nabla} F_\tau(U_\tau)\right\|_{(\overrightarrow{\nabla}^2 F_\tau(U_\tau))^{-1}}\,. \tag{17}$$

Therefore using Lemma 19 we have that

$$\overrightarrow{\nabla}^2 \hat{F}_\tau(U_\tau) \leq \left(\frac{19}{14}\right)^2 \overrightarrow{\nabla}^2 \hat{F}_\tau(U^\star_{\tau+1})\,.$$

By the conditions on the solver, we have

$$\left\|\overrightarrow{\nabla} F_\tau(U_{\tau+1})\right\|_{(\nabla^2 F_\tau(U_{\tau+1}))^{-1}} \leq \frac{1}{8\sqrt{\eta}}\,.$$

This implies using (17) and Lemma 19 that

$$\nabla^2 \hat{F}_\tau(U^\star_{\tau+1}) \leq \left(\frac{7}{6}\right)^2 \nabla^2 \hat{F}_\tau(U_{\tau+1})\,.$$

Combining everything yields

$$\min_{\Xi\in[U_\tau, U_{\tau+1}]} \left\|\overrightarrow{\nabla} F_\tau(U_\tau) - \overrightarrow{\nabla} F_\tau(U_{\tau+1})\right\|_{[\nabla^2 F_\tau(\Xi)]^{-1}} \leq \frac{5}{24\sqrt{\eta}} + \frac{1}{8\sqrt{\eta}} \cdot \frac{7}{6} \cdot \frac{19}{14} \leq \frac{1}{2\sqrt{\eta}}\,.$$

Finally chaining Lemma 18 and 27 yields

$$\lambda^{-1} U_{\tau+1} \preceq U_\tau \preceq \lambda U_{\tau+1}$$

for

$$\lambda = 1 + \frac{\frac{1}{2}}{1 - \frac{1}{2}} \leq 2\,.$$

Further chaining (16), (17), Lemma 18 and 27 similarly as above implies $U^\star_{\tau+1} \preceq 2U_\tau$

■

Finally we provide some loose upper bounds on the inverses of the iterates.

Lemma 35 *If η and β satisfy constraints (10) and (11), we have that for any τ such that the reset condition is not triggered upto index $\tau - 1$,*

$$U_\tau^{-1} \preceq 2\left(\frac{(1+\beta)^2 T}{4\beta}\eta + d\right)\mathbf{I}_d$$
$$\text{and } X_\tau^{-1} \preceq P_\tau^{-1} \preceq \left(\frac{(1+\beta)^2 T}{4\beta^2}\eta + \frac{d}{\beta})\right)\mathbf{I}_d.$$

If further T satisfies constraint (12), then the last term is upper bounded by $T^2\mathbf{I}_d$.

Proof Lemma 22 shows that for all Hermitian matrices H, such that $\mathrm{Tr}(H) = 0$ we have that $\langle \nabla \hat{F}_{\tau-1}(U_\tau^*), H\rangle = 0$. Further since $\nabla \hat{F}_{\tau-1}(U_\tau^*)$ is Hermitian, these facts imply that $\nabla \hat{F}_{\tau-1}(U_\tau^*) = \gamma \mathbf{I}_d$ for some $\gamma \in \mathbb{R}$. Substituting the definition of $\nabla \hat{F}_{\tau-1}$ we get that

$$\gamma \mathbf{I}_d = \sum_{s=1}^{\tau-1}\left(-\frac{(1+\beta)R_s}{\langle X_s, R_s\rangle} + \beta\frac{\langle U_\tau^*, R_s\rangle R_s}{\langle X_s, R_s\rangle^2}\right) - \eta^{-1}U_\tau^{\star-1}. \tag{18}$$

Using Lemma 31 we get that $\langle U_\tau^\star, R_s\rangle \le \frac{\langle X_s, R_s\rangle}{\beta}$ and therefore the above equality implies that

$$\eta^{-1}U_\tau^{\star-1} \preceq -\gamma\mathbf{I}_d.$$

Further using (18) we have that

$$\begin{aligned}
\gamma = \langle U_\tau^\star, \nabla\hat{F}_{\tau-1}(U_\tau^\star)\rangle &= \langle U_\tau^\star, \left(\sum_{s=\tau_{i-1}}^{\tau-1}\left(-\frac{(1+\beta)R_s}{\langle X_s, R_s\rangle} + \beta\frac{\langle U_\tau^\star, R_s\rangle R_s}{\langle X_s, R_s\rangle^2}\right) - \eta^{-1}U_\tau^{\star-1}\right)\rangle \\
&= \sum_{s=\tau_{i-1}}^{\tau-1}\left(-\frac{(1+\beta)\langle U_\tau^\star, R_s\rangle}{\langle X_s, R_s\rangle} + \beta\frac{\langle U_\tau^\star, R_s\rangle^2}{\langle X_s, R_s\rangle^2}\right) - \frac{d}{\eta} \\
&\ge -\frac{(1+\beta)^2 T}{4\beta} - \frac{d}{\eta}.
\end{aligned}$$

Combining these leads to

$$U_\tau^{\star-1} \preceq (\frac{(1+\beta)^2 T}{4\beta}\eta + d)\mathbf{I}_d$$

By Lemma 31, we have $U_\tau^{-1} \preceq 2U_\tau^{\star-1}$ and by the reset condition we have $P_\tau \preceq \frac{1}{2\beta}U_\tau^{-1}$, which completes the first part of the lemma. Finally, since $\eta \le \beta/4$, $(1+\beta)^2 \le 3$ and $T \ge \max(2d, \beta^{-1})$, we have

$$\frac{(1+\beta)^2 T}{4\beta^2}\eta + \frac{d}{\beta} \le T^2\,.$$

■

E.4. Main proofs

For the next three lemmas, once again for brevity we drop the epoch superscript. Further we define the inherent dimension of the problem $\tilde{d}$ as d for the standard optimal portfolio case and d^2 for the quantum case. The next lemma bounds the cost of bias in our algorithm.

Lemma 36 *Let η, β, T satisfy constraints (10)-(12). Consider any epoch e with the reset points $\mathcal{T}_{e-1} < \mathcal{T}_e \leq T$. Let L represent the length of the epoch, i.e. $L = \mathcal{T}_e - \mathcal{T}_{e-1}$. Then the cost of bias within the epoch is bounded as follows*

$$\sum_{\tau=1}^{L} \langle X_\tau, P_\tau - P_{\tau-1} \rangle = \sum_{\tau=1}^{L} \langle P_\tau^{-1}, P_\tau - P_{\tau-1} \rangle \leq 2d \log(T)\,.$$

Proof By using Lemma 23 and Lemma 29, we have

$$\sum_{\tau=1}^{L} \langle X_\tau, P_\tau - P_{\tau-1} \rangle = \sum_{\tau=1}^{L} \langle P_\tau^{-1}, P_\tau - P_{\tau-1} \rangle \leq \log\det(P_L/d)\,.$$

Finally by Lemma 35, we have $P_L \preceq T^2 \mathbf{I}_d$, which completes the proof. ∎

Lemma 37 *For all τ, the approximation gap $\hat{F}_\tau(U_{\tau+1}) - \hat{F}_\tau(U^\star_{\tau+1})$ is bounded by*

$$\hat{F}_\tau(U_{\tau+1}) - \hat{F}_\tau(U^\star_{\tau+1}) \leq \frac{1}{56\eta}\,.$$

Proof By convexity, Cauchy-Schwartz and Lemma 18, we have

$$\begin{aligned}
\hat{F}_\tau(U_{\tau+1}) - \hat{F}_\tau(U^\star_{\tau+1}) &\leq \langle \nabla \hat{F}_\tau(U_{\tau+1}), U_{\tau+1} - U^\star_{\tau+1} \rangle \\
&\leq \left\| \vec{\nabla} \hat{F}_\tau(U_{\tau+1}) \right\|_{[\vec{\nabla}^2 \hat{F}_\tau(U_{\tau+1})]^{-1}} \left\| \vec{U}_{\tau+1} - \vec{U}^\star_{\tau+1} \right\|_{\vec{\nabla}^2 \hat{F}_\tau(U_{\tau+1})} \\
&\leq \frac{\left\| \vec{\nabla} \hat{F}_\tau(U_{\tau+1}) \right\|^2_{[\vec{\nabla}^2 \hat{F}_\tau(U_{\tau+1})]^{-1}}}{1 - \sqrt{\eta} \left\| \vec{\nabla} \hat{F}_\tau(U_{\tau+1}) \right\|_{[\vec{\nabla}^2 \hat{F}_\tau(U_{\tau+1})]^{-1}}}\,.
\end{aligned}$$

The algorithm requires

$$\left\| \vec{\nabla} \hat{F}_\tau(U_{\tau+1}) \right\|_{[\nabla^2 \hat{F}_\tau(U_{\tau+1})]^{-1}} \leq \frac{1}{8\sqrt{\eta}}\,,$$

which completes the proof. ∎

The following lemma bounds the regret with respect to biased surrogate functions g_τ within an epoch.

Lemma 38 *Let η, β, T satisfy constraints (10)-(12). Consider any epoch e with the reset points $\mathcal{T}_{e-1} < \mathcal{T}_e \leq T$. Let L represent the length of the epoch, i.e. $L = \mathcal{T}_e - \mathcal{T}_{e-1}$. The FTRL-regret with respect to any comparator U over the functions $(g_\tau)_{\tau=1}^L$ is bounded by*

$$\sum_{\tau=1}^{L} (g_\tau(X_\tau) - g_\tau(U)) \leq \frac{11\tilde{d}}{\beta} \log(T) + \eta^{-1} R(U)\,,$$

where $\tilde{d}$ is the inherent dimension of the problem: d^2 for the PSD case and d for the simplex case.

Proof We wish to apply Lemma 47. We first check the conditions. Note that via (13) and Lemma 27 we see that Assumption 3 is satisfied with $c_1 = (1+6\eta)^2$. Further the condition on the solver implies that Lemma 47 holds with with factor $\frac{c_1}{(1-\delta)^2} \le \frac{(1+6\eta)^2}{(1-6\eta)^2} \le \frac{6}{5}$. Therefore using Lemma 47 we have that

$$\sum_{\tau=1}^{L}(g_\tau(X_\tau) - g_\tau(U)) - \eta^{-1}R(U) \le \frac{3}{5}\sum_{\tau=1}^{L}\left\|\vec{\nabla} g_\tau(X_\tau)\right\|^2_{(\vec{\nabla}^2 G_\tau(X_\tau))^{-1}}$$
$$\le \sum_{\tau=1}^{L}\left(\left\|\vec{\nabla}\hat{f}_\tau(X_\tau)\right\|^2_{(\vec{\nabla}^2 G_\tau(X_\tau))^{-1}} + \frac{3}{2}B^2\left\|\vec{P}_\tau - \vec{P}_{\tau-1}\right\|^2_{(\vec{\nabla}^2 G_\tau(X_\tau))^{-1}}\right),$$

where we used $(a+b)^2 \le \lambda a^2 + \frac{\lambda}{\lambda-1}b^2$ for any $\lambda > 1$, generalizing it appropriately to vectors. We deal with the above two terms separately. To control the first term note the following

$$\vec{\nabla}^2 G_\tau(X_\tau) = \eta^{-1}\vec{\nabla}^2 R(X_\tau) + \sum_{s=1}^{\tau}\frac{\beta}{2}\vec{\nabla}\hat{f}_s(X_s)\vec{\nabla}\hat{f}_s(X_s)^\top.$$

Using Lemma 20 we get that for any τ

$$\left\|\vec{\nabla}\hat{f}_\tau(X_\tau)\right\|^2_{(\vec{\nabla}^2 G_\tau(X_\tau))^{-1}} \le \langle\vec{\nabla}\hat{f}_\tau(X_\tau)[\vec{\nabla}\hat{f}_\tau(X_\tau)]^\top, \left(\eta^{-1}\mathbf{I}_{\tilde{d}} + \sum_{s=1}^{\tau}\frac{\beta}{2}\vec{\nabla}\hat{f}_s(X_s)\vec{\nabla}\hat{f}_s(X_s)^\top\right)^{-1}\rangle.$$

Using Lemma 23, the following computation follows.

$$\sum_{\tau=1}^{L}\left\|\nabla\hat{f}_\tau(X_\tau)\right\|^2_{(\vec{\nabla}^2 G_\tau(X_t))^{-1}} \le \frac{2}{\beta}\log\det(\mathbf{I}_{\tilde{d}} + \sum_{\tau=1}^{L}\frac{\eta\beta}{2}\vec{\nabla}\hat{f}_\tau(X_\tau)\vec{\nabla}\hat{f}_\tau(X_\tau)^\top)$$
$$\le \frac{2}{\beta}\tilde{d}\log\left((\tilde{d} + \frac{\eta\beta}{2}T\max_{t\in[\tau]}\left\|\vec{\nabla}\hat{f}_\tau(X_\tau)\right\|^2)/\tilde{d}\right).$$

By Lemma 35, we have $X_\tau \succeq T^{-2}\mathbf{I}_d$ for all $\tau \in [L]$, hence

$$\left\|\vec{\nabla}\hat{f}_\tau(X_\tau)\right\|^2 = \frac{\langle R_\tau, R_\tau\rangle}{\langle X_\tau, R_\tau\rangle^2} \le T^4.$$

Further, since $T > 2d$ and $\eta \le \frac{1}{2}$, we have

$$\sum_{\tau=1}^{L}\left\|\vec{\nabla}\hat{f}_\tau(X_\tau)\right\|^2_{(\vec{\nabla}^2 G_\tau(X_t))^{-1}} \le \frac{10}{\beta}\tilde{d}\log(T).$$

For the second norm, we have

$$\begin{aligned}
\sum_{t=1}^{\tau}\left\|\overrightarrow{P}_\tau-\overrightarrow{P}_{\tau-1}\right\|^2_{(\nabla^2 G_\tau(X_t))^{-1}} &\leq \eta\sum_{t=1}^{\tau}\left\|\overrightarrow{P}_\tau-\overrightarrow{P}_{\tau-1}\right\|^2_{(\nabla^2 R(X_\tau))^{-1}} \\
&= \eta\sum_{t=1}^{\tau}\|P_\tau-P_{\tau-1}\|^2_{X_\tau} \quad \text{(by Lemma 20)} \\
&= \eta\sum_{t=1}^{\tau}\operatorname{Tr}\left((\mathbf{I}_d - X_\tau^{\frac{1}{2}}P_{\tau-1}X_\tau^{\frac{1}{2}})_+^2\right) \\
&\leq \eta\sum_{t=1}^{\tau}\operatorname{Tr}\left((\mathbf{I}_d - X_\tau^{\frac{1}{2}}P_{\tau-1}X_\tau^{\frac{1}{2}})_+(\mathbf{I}_d - X_\tau^{\frac{1}{2}}X_{\tau-1}^{-1}X_\tau^{\frac{1}{2}})_+\right) \\
&\leq 6\eta^2\sum_{t=1}^{\tau}\operatorname{Tr}\left((\mathbf{I}_d - X_\tau^{\frac{1}{2}}P_{\tau-1}X_\tau^{\frac{1}{2}})_+\right) \\
&= 6\eta^2\sum_{t=1}^{\tau}\langle X_\tau, X_\tau^{-1/2}(\mathbf{I}_d - X_\tau^{\frac{1}{2}}P_{\tau-1}X_\tau^{\frac{1}{2}})_+X_\tau^{-1/2}\rangle \\
&= 6\eta^2\sum_{t=1}^{\tau}\langle X_\tau, P_\tau - P_{\tau-1}\rangle \\
&= 6\eta^2\sum_{t=1}^{\tau}\langle P_\tau^{-1}, P_\tau - P_{\tau-1}\rangle \leq 12\eta^2 d\log(T)\,,
\end{aligned}$$

where we use $X_\tau^{\frac{1}{2}}X_{\tau-1}^{-1}X_\tau^{\frac{1}{2}} \succeq \frac{1}{1+6\eta}\mathbf{I}_d$ by Lemma 33 for the third inequality and Lemma 36 for the last equality. By $\eta \leq \frac{1}{4B}$, $\beta \leq \sqrt{2}-1$ by constraint (10),(11), we have

$$\frac{3}{2}B^2\sum_{t=1}^{\tau}\left\|\overrightarrow{P}_\tau-\overrightarrow{P}_{t-1}\right\|^2_{(\overrightarrow{\nabla}^2 G_\tau(X_t))^{-1}} \leq \frac{36d}{32}\log(T) \leq \frac{\tilde{d}}{\beta}\log(T).$$

Combining both bounds completes the proof. ∎

The following lemma lower bounds the negative regret contribution we get.

Lemma 39 *If η, β satisfy constraints (10) and (11), then for any τ, the negative regret is bounded by*

$$-\langle U_{\tau+1}, P_\tau - P_0\rangle B \leq \mathbb{I}\{\textit{reset happened at } \tau\}\left(-\frac{5B}{12\beta}+dB\right).$$

Proof If no reset happened at τ, we have $P_\tau \succeq P_0$ and the term is bounded by 0. Otherwise by Lemma 33 and the reset condition, we have

$$\langle U_{\tau+1}, P_\tau\rangle \geq \frac{1}{1+6\eta}\langle U_{\tau+1}, P_{\tau+1}\rangle \geq \frac{1}{2(1+6\eta)^2\beta}\,.$$

By the constraint (10), we have $(1+6\eta)^2 \leq \frac{6}{5}$. Using $P_0 = d\mathbf{I}_d$ completes the proof. ∎

Proof of Theorem 1 and Theorem 5. We use $\tilde{d}$ to denote d^2 in the full PSD case and d in the regular portfolio case (i.e. all matrices are diagonal matrices). With

$$B = \frac{264}{5}\tilde{d}\log(T)$$
$$\beta = \frac{11\tilde{d}}{7Bd}$$
$$\eta = \frac{1}{4B}\,,$$

the constraints can be seen to (10)-(12) be satisfied. Consider any epoch e with the reset points $\mathcal{T}_{e-1} < \mathcal{T}_e \leq T$. Let L represent the length of the epoch, i.e. $L = \mathcal{T}_e - \mathcal{T}_{e-1}$. We drop the superscript e below for brevity. Then for any comparator $U \succeq T^{-1}\mathbf{I}_d$, we have that

$$\begin{aligned}
\mathrm{Reg}_e(U) &= \sum_{t=\mathcal{T}_{e-1}}^{\mathcal{T}_e-1}(f_t(X_t) - f_t(U)) \leq \sum_{\tau=1}^{L}(\underline{\hat{f}_\tau}(X_\tau) - \underline{\hat{f}_\tau}(U)) && \text{(by Lemma 7)}\\
&\leq \max_{U'\in\mathcal{A}}\sum_{t=1}^{\tau}(\underline{\hat{f}_\tau}(X_\tau) - \underline{\hat{f}_\tau}(U')) - \eta^{-1}R(U') + \eta^{-1}R(U)\\
&= \sum_{\tau=1}^{L}(\hat{f}_\tau(X_\tau) - \hat{f}_\tau(U^\star_{L+1})) - \eta^{-1}R(U^\star_{L+1}) + \eta^{-1}R(U) && \text{(by Lemma 8)}\\
&= \sum_{\tau=1}^{L}(g_\tau(X_\tau) - g_\tau(U_{L+1})) - \eta^{-1}R(U_{L+1}) + \sum_{t=1}^{\tau}\langle X_\tau - U_{L+1}, B(P_\tau - P_{t-1})\rangle\\
&\qquad + \hat{F}_L(U_{L+1}) - \hat{F}_L(U^\star_{L+1}) + \underbrace{\eta^{-1}R(U)}_{\leq \eta^{-1}d\log(T)}\,,\\
&\leq \frac{11}{\beta}\tilde{d}\log(T) + \frac{1}{56\eta} + 2d\log(T)B + \frac{d\log(T)}{\eta} - (\frac{5B}{12\beta} - dB)\mathbb{I}\{\text{reset happened at }\tau\}\\
& && \text{(by Lemma 36-39)}\\
&\leq \frac{11}{\beta}\tilde{d}\log(T) + 7d\log(T)B - \frac{5B}{12\beta}\mathbb{I}\{\text{reset happened at }\tau\}\\
&= \frac{3696}{5}d\tilde{d}\log^2(T) - \frac{3696}{5}d\tilde{d}\log^2(T)\mathbb{I}\{\text{reset happened at }\tau\}\,.
\end{aligned}$$

■

Proof of Corollary 2 and Corollary 6. We use $\tilde{d}$ to denote d^2 in the full PSD case and d in the regular portfolio case (i.e. all matrices are diagonal matrices). Define $U^\circ = \arg\min_{U\in\mathcal{A}}\sum_{t=1}^{T} f_t(X)$ and $U = (1 - \frac{d}{T})U^\circ + \frac{d}{T}(\frac{1}{d}\mathbf{I}_d)$. By construction $U \succeq T^{-1}\mathbf{I}_d$ is satisfied. As denoted earlier $\mathcal{T}_1, \ldots, \mathcal{T}_E$ are the reset points of Algorithm 3 over the game with T_0 steps, and $\mathcal{T}_0 = 1$ and $\mathcal{T}_{E+1} = T + 1$ by convention. We now derive the following succession of inequalities

$$\begin{aligned}
\mathrm{Reg} &\leq \mathrm{Reg}(U) - T\log\left(1 - \frac{d}{T}\right) \leq \mathrm{Reg}(U) + \mathcal{O}(d)\\
&\leq \sum_{e=0}^{E}\sum_{t\in\mathcal{E}_e}(f_t(X_t) - f_t(U)) + \mathcal{O}(d) \leq \mathcal{O}(d\tilde{d}\log^2(T))\,,
\end{aligned}$$

where the first inequality follows via a simple bound on the optimality gap between U° and U and the last step uses the epoch-wise regret bounds established in Theorem 1 and Theorem 5. ■

Proof of Lemma 8 In the proof we omit the superscript e for brevity. Define for any s, the set $\mathcal{D}_s := \{X \in \mathcal{A} | \langle X, R_s\rangle \leq \beta^{-1}\langle X_s, R_s\rangle\}$. As we have shown in Section 4 we have that for all s, $\underline{\hat{f}_s}|_{\mathcal{D}_s} = \hat{f}_s|_{\mathcal{D}_s}$ where $l|_S$ for a function l and a set S denotes the restriction of the function l on the set S. The first step is to show the following for any step τ

$$\underline{D_\tau} := \{U \in \mathcal{A} \,|\, U \preceq \beta^{-1}P_\tau^{-1}\} \subset \bigcap_{s=1}^{\tau} \mathcal{D}_s\,.$$

To derive the above, note that due to $U \in \underline{D_\tau}$, considering any $s \leq \tau$ and noting that $R_s \succeq 0$, we have

$$\begin{aligned}\frac{\langle U, R_s\rangle}{\langle X_s, R_s\rangle} &\leq \sup_{R\in\mathcal{H}_+^d} \frac{\langle U, R\rangle}{\langle X_s, R\rangle} = \sup_{R'\in\mathcal{H}_+^d} \frac{\langle U, X_s^{-\frac{1}{2}}R'X_s^{-\frac{1}{2}}\rangle}{\langle X_s, X_s^{-\frac{1}{2}}R'X_s^{-\frac{1}{2}}\rangle} = \sup_{R'\in\mathcal{H}_+^d} \frac{\langle X_s^{-\frac{1}{2}}UX_s^{-\frac{1}{2}}, R'\rangle}{\mathrm{Tr}(R')}\\ &= \max_i \mathrm{ev}_i(X_s^{-\frac{1}{2}}UX_s^{-\frac{1}{2}}) \leq \max_i \mathrm{ev}_i(P_\tau^{\frac{1}{2}}UP_\tau^{\frac{1}{2}}) \leq \beta^{-1}\,,\end{aligned}$$

which concludes that claim. Next we show that $U^\star_{L+1} \in \mathrm{int}(\underline{D_L})$. Since $L-1$ did not trigger a reset, we know that $U_L \prec \frac{1}{2(1+6\eta)\beta}P_L^{-1}$. By Lemma 33 and 34, we have $U^\star_{L+1} \preceq 2U_L$ and $P_L^{-1} \preceq (1+6\eta)P_{L+1}^{-1}$. Hence $U^\star_{L+1} \prec \beta^{-1}P_{\tau+1}^{-1}$. Finally since $\underline{\hat{f}_t}|_{\underline{D_t}} = \hat{f}_t|_{\underline{D_t}}$ and $U^\star_{L+1}$ is by definition the minimizer

$$U^\star_{L+1} = \arg\min_{X\in\mathcal{A}} \sum_{s=1}^{\tau} \hat{f}_t(X) + \eta^{-1}R(X)\,,$$

this implies that $U^\star_{L+1}$ is a local minimum and by convexity a global minimum of the LHS in Lemma 8. ■

Appendix F. Solving the SCHRÖDINGER'S-BISONS optimization problem

Proof [Proof of Lemma 3] We drop the superscript notation for e for brevity through this proof. We will prove the statement by induction. The inductive assumption along with (13), gives us that

$$\left\|\overrightarrow{\nabla} G_\tau(X_\tau)\right\|_{[\nabla^2 G_\tau(X_\tau)]^{-1}} \leq 3\sqrt{\eta}\,.$$

Further by Lemma 21 the functions $\frac{1}{\eta}R, G_\tau, \hat{F}_\tau$ are $\sqrt{\eta}$ self-concordant. Let $X_{\tau+1} = X^+$ be the one step damped Newton update defined in Lemma 15. Therefore by Lemma 15 we have that:

$$\left\|\overrightarrow{\nabla} G_\tau(X_{\tau+1})\right\|_{[\nabla^2 G_\tau(X_{\tau+1})]^{-1}} \leq 2\sqrt{\eta}\left\|\overrightarrow{\nabla} G_\tau(X_\tau)\right\|^2_{[\nabla^2 G_\tau(X_\tau)]^{-1}} \leq 6\eta\left\|\overrightarrow{\nabla} G_\tau(X_\tau)\right\|_{[\nabla^2 G_\tau(X_\tau)]^{-1}} \leq 18\eta^{3/2} \leq 6\eta\,.$$

Similarly for U, we have by (16)

$$\left\|\overrightarrow{\nabla} \hat{F}_\tau(U_\tau)\right\|_{[\nabla^2 \hat{F}_\tau(U_\tau)]^{-1}} \leq \frac{5}{24\sqrt{\eta}}\,.$$

Let $U_{\tau+1} = U^+$ be the one step damped Newton update defined in Lemma 15. Therefore by Lemma 15 we have that:

$$\left\| \overrightarrow{\nabla} \hat{F}_\tau(U_{\tau+1}) \right\|_{[\nabla^2 \hat{F}_\tau(U_{\tau+1})]^{-1}} \le 2\sqrt{\eta} \left\| \overrightarrow{\nabla} \hat{F}_\tau(U_\tau) \right\|^2_{[\nabla^2 \hat{F}_\tau(U_\tau)]^{-1}} \le \frac{1}{8\sqrt{\eta}}\,.$$

■

Appendix G. FTRL lower bound omitted proofs.

First, we prove Lemma 10. This lemma follows from Lemma 48, since $\Pi\mathcal{A}$ has non-zero volume, and the fact that Assumption 4 holds, as shown by the following lemma:

Lemma 40 *For any $\eta > 0$,* LB-FTRL *satisfies Assumption 4 with $c_2 = \frac{1}{(1+\eta)^2}$.*

Proof First note that for any $x, y \in \mathrm{int}(\Delta([d]))$, we have

$$\nabla^2 F_t(x) = \sum_{s=1}^{t} \frac{(\Pi r_s)(\Pi r_s)^\top}{\langle x, r_s\rangle^2} + \sum_{i=1}^{d} \frac{(\Pi \mathbf{e}_i)(\Pi \mathbf{e}_i)^\top}{\langle x, \mathbf{e}_i\rangle^2} \succeq \min_{i\in[d]} \frac{y_i^2}{x_i^2} \nabla^2 F_t(y)\,.$$

Hence we need to prove that for $c_2 = \frac{1}{(1+\eta)^2}$, we have $\min_{i\in[d]} x_{t,i}/x^\lambda_{t,i} \ge \sqrt{c_2}$.

We have

$$\begin{aligned} D_{G_t^\lambda}(x_t^\lambda, x_t) + D_{G_t^\lambda}(x_t, x_t^\lambda) &= \langle x_t^\lambda - x_t, \nabla G_t^\lambda(x_t^\lambda) - \nabla G_t^\lambda(x_t)\rangle \\ &= \langle x_t^\lambda - x_t, -\lambda \nabla f_t(x_t)\rangle \\ &= \lambda\left(\frac{\langle x_t^\lambda, r_t\rangle}{\langle x_t, r_t\rangle} - 1\right)\,. \end{aligned}$$

Let $H_t^\lambda(x) = G_t^\lambda - \sum_{s=1}^{t-1} f_s(x)$, then

$$\begin{aligned} D_{H_t^\lambda}(x_t^\lambda, x_t) + D_{H_t^\lambda}(x_t, x_t^\lambda) &= \langle x_t^\lambda - x_t, \nabla H_t^\lambda(x_t^\lambda) - \nabla H_t^\lambda(x_t)\rangle \\ &= \eta^{-1}\sum_{i=1}^{d}\left(\frac{x^\lambda_{t,i}}{x_{t,i}} + \frac{x_{t,i}}{x^\lambda_{t,i}} - 2\right) + \lambda\left(\frac{\langle x_t^\lambda, r_t\rangle}{\langle x_t, r_t\rangle} + \frac{\langle x_t, r_t\rangle}{\langle x_t^\lambda, r_t\rangle} - 2\right) \end{aligned}$$

Since by construction $\nabla^2 G_t^\lambda \succeq \nabla^2 H_t^\lambda$, we have

$$D_{G_t^\lambda}(x_t^\lambda, x_t) + D_{G_t^\lambda}(x_t, x_t^\lambda) \ge D_{H_t^\lambda}(x_t^\lambda, x_t) + D_{H_t^\lambda}(x_t, x_t^\lambda)\,,$$

which implies

$$\lambda\left(1 - \frac{\langle x_t, r_t\rangle}{\langle x_t^\lambda, r_t\rangle}\right) \ge \eta^{-1}\sum_{i=1}^{d}\left(\frac{x^\lambda_{t,i}}{x_{t,i}} + \frac{x_{t,i}}{x^\lambda_{t,i}} - 2\right)\,.$$

Let $z = \min_{i\in[d]} \frac{x_{t,i}}{x^\lambda_{t,i}}$, then this results in

$$\begin{aligned} & (1-z) \ge \eta^{-1}(z^{-1} - z - 2) \\ \Leftrightarrow \qquad & z \ge \frac{1}{1+\eta}\,, \end{aligned}$$

as required. ■

In the remainder of this section, we use $f(d,T) = \mathcal{O}(g(d,T))$, $f(d,T) = \Omega(g(d,T))$ to mean that there exists universal constants $C > c > 0$ and $T_0 = \text{Poly}(d)$ such that for all $T > T_0$, it holds $f(d,T) \leq Cg(d,T)$ and $f(d,T) \geq cg(d,T)$ respectively. Poly(x) hereby means that there exists some fixed exponent $a \in [0,\infty)$ such that the statement holds for x^a. Finally $f(d,T) = \Theta(g(d,T))$ means $f(d,T) = \mathcal{O}(g(d,T))$ and $f(d,T) = \Omega(g(d,T))$ hold simultaneously. Also recall that we assume $T > T_0 = \text{Poly}(\mathcal{T}, d)$, specifically we will use $\mathcal{T} \leq T^\alpha$ throughout this section.

Define the scaling factors $(c_i)_{i=0}^I = 1 - 2^i T^{-\alpha}$, where $I = \lfloor \frac{1}{3}\log_2(T^\alpha)\rfloor$. For $\boldsymbol{x} \in \Delta([d])$, we define the "pulling to the center" operator $^{(s)}$, by $\boldsymbol{x}^{(s)} = \Pi^{-1}(c_s \Pi \boldsymbol{x}) = c_s \boldsymbol{x} + (1-c_s)\boldsymbol{c}$.

Algorithm 4: Sequence for large regret.

```
Input: (t_i, o_i)_{i=1}^{𝒯}, α = 1/8, T
for i = 1, 2, ..., 𝒯 do
    for k = 1, ..., T^α do
        for s = 1, ..., ⌊(1/3) α log_2(T)⌋ do
            while x_t ≠ t_i^(s) do
                r_t ← move-to-x(t_i^(s); F_{t−1})
                t ← t + 1
            end
            r_t ← o_i^(s)
            t ← t + 1
        end
    end
end
Function move-to-x(x ; F):
    g ← Π∇F(x)
    g ← min{T^{−1/2} / ‖g‖_2 , 1/(d max{1−⟨g,Πx⟩,0})} g
    return: Π^{−1} g
```

Basic calculations: By definition $c_i = \Theta(1)$ for all $i \in [I] \cup \{0\}$. Further we have for any $s, s' \in [I] \cup \{0\}$:

$$1 - c_s c_{s'} = (2^s + 2^{s'})T^{-\alpha} - 2^{s+s'}t^{-2\alpha} = \Theta(2^{\max\{s,s'\}}T^{-\alpha})\,.$$

For any $\boldsymbol{x}, \boldsymbol{y} \in \Delta([d])$, we have

$$\langle \boldsymbol{x}^{(s)}, \boldsymbol{y}^{(s')}\rangle = \frac{1}{d} + \langle \Pi\boldsymbol{x}^{(s)}, \Pi\boldsymbol{y}^{(s')}\rangle = \frac{1}{d} + c_s c_{s'}\langle \Pi\boldsymbol{x}, \Pi\boldsymbol{y}\rangle = \frac{1 - c_s c_{s'}}{d} + c_s c_{s'}\langle \boldsymbol{x}, \boldsymbol{y}\rangle\,.$$

By the assumption on the sequence, we have for any $j < i$:

$$\langle \boldsymbol{t}_i^{(s)}, \boldsymbol{o}_j^{(s')}\rangle = \Omega(\langle \boldsymbol{t}_i, \boldsymbol{o}\rangle) = \Omega\left(\frac{1}{\text{Poly}(d)}\right) \tag{19}$$

$$\langle \boldsymbol{t}_i^{(s)}, \boldsymbol{o}_i^{(s')}\rangle = \frac{1 - c_s c_{s'}}{d} = \Theta\left(\frac{2^{\max\{s,s'\}}}{dT^\alpha}\right) \tag{20}$$

Bounding the movement steps. The main result of this section is the following Lemma.

Lemma 41 *The number of movement steps up to time T is bounded by $\mathcal{O}\left(\textit{Poly}(d)T^{3\alpha+\frac{1}{2}}\log(T)^2\right)$.*

In order to prove this Lemma, we first require the following.

Lemma 42 *The while routine over move-to-x for a target $\boldsymbol{t}$ up from time t requires $\tau \leq \frac{2T^{1/2}}{d}\left\|\nabla_\Pi F_t(\boldsymbol{t})\right\|_2 + 1$ steps.*

Proof We have reached the target, if at time $t+\tau$ is holds $\nabla_\Pi F_{t+\tau}(\boldsymbol{t}) = 0$. We select the movement returns r_s for $s \in \{t, \ldots, t+\tau-1\}$ such that

$$\left\|\nabla_\Pi F_{s+1}(\boldsymbol{t})\right\|_2 = \max\{0, \left\|\nabla_\Pi F_s(\boldsymbol{t})\right\|_2 - \left\|\nabla_\Pi f_{s+1}(\boldsymbol{t})\right\|_2\}\,.$$

When we cannot reach the target in one step, the norm of the gradient is

$$\left\|\nabla_\Pi f_{s+1}(\boldsymbol{t})\right\|_2 = \frac{\|\Pi r_s\|}{\frac{1}{d} + \langle \Pi \boldsymbol{t}, \Pi r_s\rangle} \geq \frac{T^{-1/2}}{\frac{1}{d} + T^{-1/2}} \geq \frac{d}{2}T^{-1/2}\,.$$

Hence the number of steps τ until the norm is 0 is bounded by

$$\tau \leq \frac{2T^{1/2}}{d}\left\|\nabla_\Pi F_t(\boldsymbol{t})\right\|_2 + 1\,.$$

■

Lemma 43 *For any movement-return r and any $x, y \in \Delta([d])$, it holds*

$$\|\nabla_\Pi f(x; r) - \nabla_\Pi f(y; r)\| = \mathcal{O}(d^2 T^{-1})\,.$$

Proof

$$\begin{aligned}\|\nabla_\Pi f(x; r) - \nabla_\Pi f(y; r)\| &= \left|\frac{1}{1/d + \langle \Pi x, \Pi r\rangle} - \frac{1}{1/d + \langle \Pi y, \Pi r\rangle}\right| \|\Pi r\| \\ &\leq \left(\frac{1}{1/d - T^{-\frac{1}{2}}} - \frac{1}{1/d + T^{-\frac{1}{2}}}\right)T^{-\frac{1}{2}} = \frac{2T^{-1}}{1/d^2 - T^{-1}} = \mathcal{O}(d^2T^{-1})\,,\end{aligned}$$

where we use that movement returns by construction satisfy $\|\Pi r\| \leq T^{-\frac{1}{2}}$ and $\|\Pi x\| \leq \|x\| \leq 1$ for any $x \in \Delta([d])$. ■

Lemma 44 *For any $x \in \Delta([d])$ and $s \in [I] \cup \{0\}$, the largest possible gradient of any regularizer part $r_i(x^{(s)}) = f(x^{(s)}; \boldsymbol{e}_i), i \in [d]$ is bounded by*

$$\max_{x \in \Delta([d])} \left\|\nabla_\Pi r_i(x^{(s)})\right\| = \mathcal{O}\left(d\frac{T^\alpha}{2^s}\right)\,.$$

Proof

$$\left\|\nabla_\Pi r_i(x^{(s)}\right\| = \left\|\frac{\Pi \boldsymbol{e}_i}{\langle x^{(s)}, \boldsymbol{e}_i\rangle}\right\| \le \frac{d}{1-c_s} = d\frac{T^\alpha}{2^s}\,,$$

where we used

$$\langle x^{(s)}, \boldsymbol{e}_i\rangle = \frac{1-c_s}{d} + c_s\langle x, \boldsymbol{e}_i\rangle \ge \frac{1-c_s}{d}\,.$$

■

Proof [Proof of Lemma 41] For the initial move-to-x, we have $F_0(\boldsymbol{t}_1^{(0)}) = R(\boldsymbol{t}_1^{(0)})$, hence by combining Lemma 44 and 42, we require $\mathcal{O}(dT^{\alpha+\frac{1}{2}})$ initial steps. Afterwards, we need to bound the steps between any two targets $\boldsymbol{t}_k^{(s)}, \boldsymbol{t}_{k'}^{(s')}$, where $k \le k'$. Assume this switch happens at time $\tau \le T$ (since the Lemma statement is concerned with movement steps before time T), directly after the agent observed a return $\boldsymbol{o}_k^{(s)}$ at target $\boldsymbol{t}_k^{(s)}$. Hence

$$\left\|\nabla_\Pi F_\tau(\boldsymbol{t}_{k'}^{(s')})\right\| \le \left\|\nabla_\Pi f(\boldsymbol{t}_{k'}^{(s')};\boldsymbol{o}_k^{(s)})\right\| + \left\|\nabla_\Pi F_{\tau-1}(\boldsymbol{t}_{k'}^{(s')}) - \nabla_\Pi F_{\tau-1}(\boldsymbol{t}_k^{(s)})\right\|\,,$$

where we use that $\left\|\nabla_\Pi F_{\tau-1}(\boldsymbol{t}_k^{(s)})\right\| = 0$ since the agent was in that point when receiving r_τ. Splitting the time-steps into movement-returns $\mathcal{M}_\tau := \{t \in [\tau]\,|\, \|\Pi r_t\| \le T^{-1/2}\}$ and regular returns yields

$$\begin{aligned}
&\left\|\nabla_\Pi f(\boldsymbol{t}_{k'}^{(s')};\boldsymbol{o}_k^{(s)})\right\| + \left\|\nabla_\Pi F_{\tau-1}(\boldsymbol{t}_{k'}^{(s')}) - \nabla_\Pi F_{\tau-1}(\boldsymbol{t}_k^{(s)})\right\| \\
&\le \left\|\nabla_\Pi R(\boldsymbol{t}_k^{(s)}) - \nabla_\Pi R(\boldsymbol{t}_{k'}^{(s')})\right\| + \left\|\sum_{s\in\mathcal{M}_\tau} \nabla_\Pi f_s(\boldsymbol{t}_k^{(s)}) - \nabla_\Pi f_s(\boldsymbol{t}_k^{(s)})\right\| \\
&\quad + T^\alpha\left(\sum_{j=1}^{k-1}\sum_{r=0}^{I}\left(\left\|\nabla_\Pi f(\boldsymbol{t}_k^{(s)};\boldsymbol{o}_j^{(r)})\right\| + \left\|\nabla_\Pi f(\boldsymbol{t}_{k'}^{(s')};\boldsymbol{o}_j^{(r)})\right\|\right)\right. \\
&\qquad \left. + \sum_{r=0}^{I}\left(\left\|\nabla_\Pi f(\boldsymbol{t}_k^{(s)};\boldsymbol{o}_k^{(r)})\right\| + \left\|\nabla_\Pi f(\boldsymbol{t}_{k'}^{(s')};\boldsymbol{o}_k^{(r)})\right\|\right)\right) \\
&\le \mathcal{O}(d^2T^\alpha) + \mathcal{O}(d^2) \qquad \text{(Lemma 44 and 43)} \\
&\quad + \mathcal{O}\left(\max_{j<\ell;r,r'\in[I]\cup\{0\}} \frac{T^\alpha \mathcal{T}\log(T)}{\langle \boldsymbol{t}_\ell^{(r)}, \boldsymbol{o}_j^{(r')}\rangle}\right) + \mathcal{O}\left(\max_{j\le\ell;r,r'\in[I]\cup\{0\}} \frac{T^\alpha \log(T)}{\langle \boldsymbol{t}_\ell^{(r)}, \boldsymbol{o}_j^{(r')}\rangle}\right) \\
&= \mathcal{O}(\mathrm{Poly}(d)T^{2\alpha}\log(T))\,. \qquad \text{(Equation (19) and (20))}
\end{aligned}$$

The proof is completed by applying Lemma 42, noting that the number of switches is bounded by $I\mathcal{T} \le T^\alpha \log(T)$. ■

Bounding the Hessian trace. We first bound the Hessian trace of movement-steps.

Lemma 45 *The movement time-steps $\mathcal{M}_\tau$ for any $\tau \le T$ and any $\boldsymbol{t} \in \Delta([d])$ satisfy*

$$\sum_{t\in\mathcal{M}_\tau} \|\nabla_\Pi f(\boldsymbol{t}; r_t)\|^2 = \mathcal{O}(d^2)\,.$$

Proof By construction $\|\Pi r_t\| \leq T^{-\frac{1}{2}}$, so

$$\|\nabla_\Pi f(\boldsymbol{t}; r_t)\|^2 = \frac{\|\Pi r_t\|^2}{(\frac{1}{d} + \langle \Pi \boldsymbol{t}, \Pi r_t\rangle)^2} \leq \frac{T^{-1}}{(\frac{1}{d} - T^{-\frac{1}{2}})^2} = \mathcal{O}(d^2 T^{-1})\,.$$

Summing over less than T time-steps completes the proof. ■

We are ready to bound the total Hessian.

Lemma 46 *Assume $\tau \leq T$ is the time-step where the m-th iteration through targets $(\boldsymbol{t}_i^{(s)})_{s=0}^I$ is completed, then the trace of the Hessian at any target $\boldsymbol{t}_i^{(s)}$ is bounded by*

$$Tr(\nabla^2_\Pi F_\tau(\boldsymbol{t}_i^{(s)})) = \mathcal{O}\left((Poly(d) + m(s+1))\, d^2 \frac{T^{2\alpha}}{2^{2s}}\, \|\Pi \boldsymbol{o}_i\|^2\right.\,.$$

Proof We split the trace into 4 terms below based on various contributions from (a) the regularizer, (b) the time steps $\mathcal{M}_\tau$ where the returns are movement-returns selected by the move-to-x subroutine, (c) the returns $\boldsymbol{o}_j^{(s)}$ selected for $j < i$ and (d) the returns selected for targets $\boldsymbol{t}_i^{(s)}$, $s \in [I] \cup \{0\}$. The first two terms are bounded by Lemma 44 and 45 respectively.

$$\begin{aligned}
\mathrm{Tr}(\nabla^2_\Pi F_\tau(\boldsymbol{t}_i^{(s)})) &= \sum_{i=1}^d \left\|\nabla_\Pi f(\boldsymbol{t}_i^{(s)}; \boldsymbol{e}_i)\right\|^2 + \sum_{s \in \mathcal{M}_\tau} \left\|\nabla_\Pi f_s(\boldsymbol{t}_i^{(s)})\right\|^2 \\
&\quad + \sum_{j=1}^{i-1}\sum_{s'=0}^{I} \left\|\nabla_\Pi f(\boldsymbol{t}_i^{(s)}; \boldsymbol{o}_j^{(s')})\right\|^2 + m \sum_{s'=0}^{I} \left\|\nabla_\Pi f(\boldsymbol{t}_i^{(s)}, \boldsymbol{o}_i^{(s')})\right\|^2 \\
&\leq d^3 \frac{T^{2\alpha}}{2^{2s}} + \mathcal{O}(d^2) + \max_{j<i, s' \in [I]\cup\{0\}} \frac{\mathcal{T} I}{\langle \boldsymbol{t}_i^{(s)}, \boldsymbol{o}_j^{(s')}\rangle^2} + m \sum_{s'=0}^{I} \frac{\|\Pi \boldsymbol{o}_i\|^2}{\langle \boldsymbol{t}_i^{(s)}, \boldsymbol{o}_i^{(s')}\rangle^2} \\
&\leq \mathcal{O}\left(d^3 \frac{T^{2\alpha}}{2^{2s}}\right) + \mathcal{O}(\mathrm{Poly}(d) T^\alpha \log(T)) \\
&\quad + m\left(\sum_{s'=0}^{s} 2^{-2s} + \sum_{s'=s+1}^{I} 2^{-2s'}\right) d^2 T^{2\alpha} \|\Pi \boldsymbol{o}_i\|^2 \\
&= \mathcal{O}\left(\mathrm{Poly}(d) \frac{T^{2\alpha}}{2^{2s}}\right) + \mathcal{O}\left(m(s+1) d^2 \frac{T^{2\alpha}}{2^{2s}}\right) \|\Pi \boldsymbol{o}_i\|^2
\end{aligned}$$

,

where we use equations (19) and (20) and the fact that $\mathcal{T} \leq T^\alpha$. $\mathcal{O}(T^\alpha \log(T) = \mathcal{O}(\frac{T^{2\alpha}}{2^{2s}})$follows from

$$2^{2s} T^{-\alpha} \log(T) \leq T^{-\alpha/3} \log(T) = \mathcal{O}(1)\,.$$

Finally, observe

$$0 = \langle \boldsymbol{t}_i, \boldsymbol{o}_i \rangle = \frac{1}{d} + \langle \Pi \boldsymbol{t}_i, \Pi \boldsymbol{o}_i \rangle \geq \frac{1}{d} - \|\Pi \boldsymbol{o}_i\|\,.$$

Hence

$$\mathcal{O}(\text{Poly}(d)\frac{T^{2\alpha}}{2^{2s}}) = \mathcal{O}(\text{Poly}(d)\frac{T^{2\alpha}}{2^{2s}}) \left\|\Pi \boldsymbol{o}_i\right\|^2 \,,$$

which concludes the proof. ■

G.1. Main lower bound proof

Proof [Proof of Theorem 12] By Lemma 41, there are $\mathcal{O}(\text{Poly}(d)T^{3\alpha+\frac{1}{2}}\log^2(T))$ movement-returns before time T and the algorithm walks through $\mathcal{O}(\mathcal{T}I) = \mathcal{O}(T^\alpha \log(T))$ regular returns, hence for $\alpha = \frac{1}{8}$, $\mathcal{O}(\text{Poly}(d)T^{7/8}\log^3(T)) = \mathcal{O}(T^{15/16})$ and there exists a sufficiently large $T_0 = \text{Poly}(d, \mathcal{T})$, such that the algorithm finishes before time T.

Next we bound the stability term. We have

$$\left\|\nabla_\Pi f_t(x_t)\right\|^2_{(\nabla^2_\Pi F_t(x_t))^{-1}} \geq \frac{\left\|\nabla_\Pi f_t(x_t)\right\|^2}{\text{Tr}(\nabla^2_\Pi F_t(x_t))}\,.$$

For the m-th time of visiting $\boldsymbol{t}_i^{(s)}$, the denominator is by Lemma 46 bounded by $\mathcal{O}((\text{Poly}(d)+m(s+1)))d^2\frac{T^{2\alpha}}{2^{2s}}\left\|\Pi \boldsymbol{o}_i\right\|^2$. For $m \geq T^{\alpha/2}$, the trace bound simplifies to $\mathcal{O}(m(s+1)d^2 2^{-2s}T^{2\alpha})\left\|\Pi \boldsymbol{o}_i\right\|^2$, since we assume $T^{\alpha/2} = \Omega(\text{Poly}(d))$. The nominator is

$$\left\|\nabla_\Pi f(\boldsymbol{t}_i^{(s)}; \boldsymbol{o}_i^{(s)})\right\|^2 = \Theta(d^2 2^{-2s}T^{2\alpha}\left\|\Pi \boldsymbol{o}_i\right\|^2)\,.$$

For the total stability, we have

$$\begin{aligned}(stab) &\geq \sum_{i=1}^{\mathcal{T}}\sum_{m=T^{\alpha/2}}^{T^\alpha}\sum_{s=0}^{I}\frac{1}{m(s+1)} \\ &= \Omega\left(\mathcal{T}\log(T)\log(I)\right)\,.\end{aligned}$$

Finally $\log(I) = \Theta(\log\log(T))$ completes the proof. ■

Appendix H. Follow-The-Regularized-Leader analysis

Both our main results rely on the standard analysis for FTRL, which we revisit in this section and extend to approximated solutions for the upper bound. Vanilla FTRL is used for online learning over a convex action set $\mathcal{X}$, where the environment picks a sequence convex loss functions $(g_t)_{t=1}^T$ from some function space $\mathcal{G}$. The input to FTRL is a regularizer $R : \mathcal{X} \to \mathbb{R}$ and define

$$x_t^\star = \arg\min_{x\in\mathcal{X}} G_{t-1}(x) := \arg\min_{x\in\mathcal{X}}\sum_{s=1}^{t-1} g_s(x) + \eta^{-1}R(x)\,.$$

The algorithm plays $x_1 = x_1^\star$ and afterwards invokes a solver $x_{t+1} = \text{APPROX-SOLVE}(G_t, x_t, \delta)$ which satisfies

$$x_{t+1} \in \{x \in \mathcal{X} \mid \left\|\nabla G_t(x)\right\|_{(\nabla^2 G_t(x))^{-1}} \leq \delta\left\|\nabla G_t(x_t)\right\|_{(\nabla^2 G_t(x_t))^{-1}}\}\,.$$

We consider in this paper special cases of FTRL that allow for a simple regret analysis.

Assumption 2 *The action set $\mathcal{X} \subset \mathbb{R}^{\tilde{d}}$ is compact and the regularizer R is strictly convex, twice continuously differentiable and goes to infinity on the boundary of $\mathcal{X}$.*

This assumption is directly satisfies by the simplex $\mathcal{A} = \Delta([d])$ and the log-barrier regularizer. Furthermore when $\mathcal{G}$ is the class of loss functions arising in the online portfolio or quantum learning problems, along with the associated log-barrier regularization, the following two assumptions are also satisfied:

Assumption 3 *There exists a universal constant c_1, such that for any sequence of functions $g_1, \dots g_T \in \mathcal{G}$ picked by the environment, any point $\bar{x}_t$ on the line between x_t and $x^\star_{t+1}$, satisfies*

$$\nabla^2 G_t(\bar{x}_t) \preceq c_1 \nabla^2 G_t(x_t)\,.$$

Assumption 4 *There exists a universal constant c_2, such that for any sequence of functions $g_1, \dots g_T \in \mathcal{G}$ picked by the environment, the interpolation between x_t and x_{t+1} defined by*

$$x_t^\lambda := \arg\min_{x \in \mathcal{X}} G_{t-1}(x) + g_t(x) - (1-\lambda)\langle x, \nabla g_t(x_t)\rangle,$$

satisfies for any $\lambda \in [0,1]$

$$\nabla^2 G_t(x_t^\lambda) \succeq c_2 \nabla^2 G_t(x_t)\,.$$

For any FTRL algorithm satisfying the assumptions above, the regret is tightly lower and upper bounded as shown in the following lemmas.

The following lemma gives an upper bound on the regret. We will prove this lemma even for the quantum case. We refer the reader to Section D for relevant definitions of gradient, Hessian and Bregman divergences in that setting.

Lemma 47 *Under Assumptions 2 and 3, the regret of δ-approximate FTRL is upper bounded for any comparator u by*

$$\sum_{t=1}^T (g_t(x_t) - g_t(u)) \le \frac{c_1}{2(1-\delta)^2} \sum_{t=1}^T \|\nabla g_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}} + \frac{R(u) - R(x_1)}{\eta}\,.$$

Proof First consider the following inequality which holds for any t,

$$\begin{aligned}
\|\nabla G_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}} &\le \left(\frac{1}{\delta}\|\nabla G_{t-1}(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}} + \frac{1}{1-\delta}\|\nabla g_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}}\right) \\
&\le \left(\delta\|\nabla G_{t-1}(x_{t-1})\|^2_{(\nabla^2 G_t(x_{t-1}))^{-1}} + \frac{1}{1-\delta}\|\nabla g_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}}\right) \\
&\le \left(\delta\|\nabla G_{t-1}(x_{t-1})\|^2_{(\nabla^2 G_{t-1}(x_{t-1}))^{-1}} + \frac{1}{1-\delta}\|\nabla g_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}}\right)
\end{aligned}$$

where the first inequality follows by noting that that $(a+b)^2 \leq \lambda a^2 + \frac{\lambda}{\lambda-1}b^2$ for any $\lambda > 1$, generalizing it appropriately to vectors and the second inequality via the approximation guarantee. This in particular implies via summation that

$$\|\nabla G_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}} \leq \frac{1}{1-\delta}\left(\sum_{\tau\leq t}\delta^{t-\tau}\|\nabla g_\tau(x_\tau)\|^2_{(\nabla^2 G_\tau(x_\tau))^{-1}}\right). \tag{21}$$

Now note that since $x^\star_{t+1}$ minimizes G_t we have that $\forall x \in \mathcal{A}: \langle x - x^\star_{t+1}, \nabla G_t(x^\star_{t+1})\rangle = 0$ (For the quantum learning case this is explicitly derived in Lemma 22). By Taylor's theorem, there exists $\lambda \in [0,1]$ such that $D_{G_t}(x^\star_{t+1}, x_t) = \frac{1}{2}\left\|x^\star_{t+1} - x_t\right\|^2_{\nabla^2 G_t(\bar{x}^\lambda_t)}$. Therefore we have that

$$\begin{aligned}
G_t(x_t) - G_t(x^\star_{t+1}) &= D_{G_t}(x_t, x^\star_{t+1}) \\
&= \langle x_t - x^\star_{t+1}, \nabla G_t(x_t) - \nabla G_t(x^\star_{t+1})\rangle - D_{G_t}(x^\star_{t+1}, x_t) \\
&= \langle x_t - x^\star_{t+1}, \nabla G_t(x_t)\rangle - \frac{1}{2}\left\|x^\star_{t+1} - x_t\right\|^2_{\nabla^2 G_t(\bar{x}^\lambda_t)} \\
&\leq \left\|x_t - x^\star_{t+1}\right\|_{\nabla^2 G_t(\bar{x}^\lambda_t)}\|\nabla G_t(x_t)\|_{\nabla^2 G_t(\bar{x}^\lambda_t)^{-1}} - \frac{1}{2}\left\|x^\star_{t+1} - x_t\right\|^2_{\nabla^2 G_t(\bar{x}^\lambda_t)} \\
&\leq \frac{1}{2}\|\nabla G_t(x_t)\|^2_{\nabla^2 G_t(\bar{x}^\lambda_t)^{-1}} \leq \frac{c_1}{2}\|\nabla G_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}}\,.
\end{aligned}$$

Summing over all time-steps and using (21) we get that

$$\sum_{t=1}^T G_t(x_t) - G_t(x^\star_{t+1}) \leq \sum_{t=1}^T \frac{c_1}{2(1-\delta)^2}\|\nabla g_t\|^2_{(\nabla^2 G_t(x_t))^{-1})}\,.$$

Next consider the decomposition below

$$\begin{aligned}
\sum_{t=1}^T (g_t(x_t) - g_t(u)) &= \sum_{t=1}^{T-1}(G_t(x_t) - G_t(x_{t+1})) + G_T(x_T) - G_T(u)) + \eta^{-1}(R(u) - R(x_1)) \\
&\leq \sum_{t=1}^T (G_t(x_t) - G_t(x^\star_{t+1})) + \underbrace{G_T(x^\star_{T+1}) - G_T(u))}_{\leq 0} + \eta^{-1}(R(u) - R(x_1))\,.
\end{aligned}$$

Combining the above conclusions we get the lemma. ■

Lemma 48 *If $\mathcal{A}$ has non-zero volume in its embedded space and Assumptions 2 and 4 are satisfied, then the regret of exact ($\delta = 0$) FTRL is lower bounded by*

$$\frac{c_2}{2}\sum_{t=1}^T \|\nabla g_t(x_t)\|^2_{(\nabla^2 G_t(x_t))^{-1}} \leq \max_{u'\in\mathcal{X}}\sum_{t=1}^T (g_t(x_t) - g_t(u'))\,.$$

Proof We have

$$\sum_{t=1}^{T}(g_t(x_t) - g_t(u)) = \sum_{t=1}^{T}(G_t(x_t) - G_t(x_{t+1})) + \underbrace{G_T(x_{T+1}) - G_T(u))}_{\leq 0} + \eta^{-1}\underbrace{(R(u) - R(x_1))}_{\geq 0} .$$

For the lower bound, we can simply lower bound $\max_{u'}$ by picking $u' = x_{T+1}$ and omit the last two terms. It remains to analyse the first term. Given that $R(x) \to \infty$ on the boundary of $\mathcal{X}$, the points x_t are all strictly in the interior of $\mathcal{A}$.

$$\begin{aligned} G_t(x_t) - G_t(x_{t+1}) &= D_{G_t}(x_t, x_{t+1}) \\ &= D_{G_t^*}(\nabla G_t(x_{t+1}), \nabla G_t(x_t)) \\ &= \frac{1}{2}\left\|\nabla G_t(x_{t+1}) - \nabla G_t(x_t)\right\|^2_{\nabla^2 G_t^*((1-\lambda)\nabla G_t(x_t) + \lambda \nabla G_t(x_{t+1}))} \\ &= \frac{1}{2}\left\|g_t(x_t)\right\|^2_{(\nabla^2 G_t(x_t^\lambda))^{-1}} \geq \frac{c_2}{2}\left\|g_t(x_t)\right\|^2_{(\nabla^2 G_t(x_t))^{-1}} . \end{aligned}$$

The above statement using the decomposition implies the lemma. ■