Data-dependent Regret Bounds for Adversarial Multi-Armed Bandits and Online Portfolio Selection

Sudeep Raja Putta

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024

Abstract

Data-dependent Regret Bounds for Adversarial Multi-Armed Bandits

and Online Portfolio Selection

Sudeep Raja Putta

This dissertation studies *Data-Dependent* regret bounds for two online learning problems. As opposed to worst-case regret bounds, data-dependent bounds are able to adapt to the particular sequence of losses seen by the player. Thus, they offer a more fine grained performance guarantee compared to worst-case bounds

We start off with the Adversarial n-Armed Bandit problem. In prior literature it was a standard practice to assume that the loss vector belonged to a known domain, typically $[0, 1]^n$ or $[-1, 1]^n$. We make no such assumption on the loss vectors, they may be completely arbitrary. We term this problem the Scale-Free Adversarial Multi Armed Bandit. At the beginning of the game, the player only knows the number of arms n. It does not know the scale and magnitude of the losses chosen by the adversary or the number of rounds T. In each round, it sees bandit feedback about the loss vectors $l_1, \ldots, l_T \in \mathbb{R}^n$. Our goal is to bound its regret as a function of n and norms of $l_1, \ldots, l_T$. We design a bandit Follow The Regularized Leader (FTRL) algorithm, that uses a log-barrier regularizer along with an adaptive learning rate tuned via the AdaFTRL technique. We give two different regret bounds, based on the exploration parameter used. With non-adaptive exploration, our algorithm has a regret of $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nT})$ and with adaptive exploration, it has a regret of $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nL_1})$. Here $L_\infty = \sup_t \|l_t\|_\infty$, $L_2 = \sum_{t=1}^T \|l_t\|_2^2$, $L_1 = \sum_{t=1}^T \|l_t\|_1$ and the $\tilde{O}$ notation suppress logarithmic factors. These are the first MAB bounds that adapt to the

$\|\cdot\|_2$, $\|\cdot\|_1$ norms of the losses. The second bound is the first data-dependent scale-free MAB bound as T does not directly appear in the regret. We also develop a new technique for obtaining a rich class of local-norm lower-bounds for Bregman Divergences. This technique plays a crucial role in our analysis for controlling the regret when using importance weighted estimators of unbounded losses.

Next, we consider the Online Portfolio Selection (OPS) problem over n assets and T time periods. This problem was first studied by Cover [1], who proposed the Universal Portfolio (UP) algorithm. UP is a computationally expensive algorithm with minimax optimal regret of $O(n \log T)$. There has been renewed interest in OPS due to a recently posed open problem Van Erven *et al.* [2] which asks for a computationally efficient algorithm that is also has minimax optimal regret. We study data-dependent regret bounds for OPS problem that adapt to the sequence of returns seen by the investor. Our proposed algorithm called AdaCurv ONS modifies the Online Newton Step(ONS) algorithm of [3] using a new adaptive curvature surrogate function for the log losses $-\log(r_t^\top w)$. We show that the AdaCurv ONS algorithm has $O(RnlogT)$ regret where R is the data-dependent quantity. For sequences where $R = O(1)$, the regret of AdaCurv ONS matches the optimal regret. However, for some sequences R could be unbounded, making the regret bound vacuous. To overcome this issue, we propose the LB-AdaCurv ONS algorithm that adds a log-barrier regularizer along with an adaptive learning rate tuned via the AdaFTRL technique. LB-AdaCurv ONS has an adaptive regret of the form $O(\min(nR \log T, \sqrt{nT \log T}))$. Thus, LB-AdaCurv ONS has a worst case regret of $O(\sqrt{nT \log T})$ while also having a data-dependent regret of $O(nR \log T)$ when $R = O(1)$. Additionally, we show logarithmic First-Order and Second-Order regret bounds for AdaCurv ONS and LB-AdaCurv ONS.

Finally, we consider the problem of Online Portfolio Selection (OPS) with predicted returns. We are the first to extend the paradigm of online learning with predictions to the portfolio selection problem. In this setting, the investor has access to noisy predictions of returns for the n assets that can be incorporated into the portfolio selection process. We propose the Optimistic Expected Utility LB-FTRL (OUE-LB-FTRL) algorithm that incorporates the predictions using a utility

function into the LB-FTRL algorithm. We explore the consistency-robustness properties for our algorithm. If the predictions are accurate, OUE-LB-FTRL's regret is $O(n \log T)$, providing a consistency guarantee. Even if the predictions are arbitrary, OUE-LB-FTRL's regret is always bounded by $O(\sqrt{nT \log T})$, providing a providing a robustness guarantee. Our algorithm also recovers a Gradual-variation regret bound for OPS. In the presence of predictions, we argue that the benchmark of static-regret becomes less meaningful. So, we consider the regret with respect to an investor who only uses predictions to select their portfolio (i.e., an expected utility investor). We provide a meta-algorithm called Best-of-Both Worlds for OPS (BoB-OPS), that combines the portfolios of an expected utility investor and a purely regret minimizing investor using a higher level portfolio selection algorithm. By instantiating the meta-algorithm and the purely regret minimizing investor with Cover's Universal Portfolio, we show that the regret of BoB-OPS with respect to the expected utility investor is $O(\log T)$. Simultaneously, BoB-OPS's static regret is $O(n \log T)$. This achieves a stronger form of consistency-robustness guarantee for OPS with predicted returns.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I want to thank my advisor, Prof. Shipra Agrawal, for her invaluable guidance over the past five years.

Dedication

Dedicated to my parents, Dr. Srinivas and Dr. Uma Devi.

Chapter 1: Introduction

This dissertation considers sequential decision-making problems within the online learning framework. Online learning enables efficient and adaptive learning in dynamic environments, making it well-suited for various applications in machine learning, artificial intelligence, reinforcement learning, and data-driven decision-making. Over the past two decades, online learning has experienced considerable progress in both theoretical understanding and practical applications, leading to its widespread adoption in the field. The importance of online learning is evident through the numerous texts, monographs, and survey articles dedicated to the subject [4, 5, 6, 7, 8, 9].

In the online learning framework, a player interacts with an environment over a series of T rounds. In each round, indexed by $t = 1, 2, \ldots, T$, the player selects w_t from a convex decision set $\mathcal{D} \subseteq \mathbb{R}^n$. The environment picks a function $f_t : \mathcal{D} \to \mathbb{R}$. Subsequently, the player incurs the cost $f_t(w_t)$ and receives feedback regarding f_t. The player's objective is to minimize the cumulative cost of interaction $\sum_{t=1}^{T} f_t(w_t)$ over the T rounds. The player's total cost can be compared to the cost of a fixed point $w \in \mathcal{D}$. This performance measure, termed as *static regret* (or just regret for short) is denoted by $\mathcal{R}_T$:

$$\mathcal{R}_T(\{f_t\}_{t=1}^T, \{w_t\}_{t=1}^T, w) = \sum_{t=1}^{T} f_t(w_t) - f_t(w)$$

In simpler terms, regret measures the difference between the player's cumulative cost and the cost of consistently using the point w. We focus on a specific category of problems within online learning, known as *online convex optimization* (OCO) [10]. In OCO, the environment is restricted to choosing only convex functions f_t. We consider two possibilities for the kind of feedback received by the player. In the *full-information* feedback setting, the player receives complete information about f_t. In the more restrictive *bandit* feedback setting, the player only receives the value

$f_t(w_t)$. A spectrum of feedback settings exists between full-information and bandit, like *semi-bandit* and *graph-structured*. However, this dissertation does not discuss these types of feedback settings.

At the beginning of each round t, the player selects the decision point w_t based on the information it has collected so far $\mathcal{I}_t$. For instance, in the full-information setting, this would be the set $\mathcal{I}_t = \{w_1, f_1, \ldots, w_{t-1}, f_{t-1}\}$. In the bandit setting, it would be $\mathcal{I}_t = \{w_1, f_1(w_1), \ldots, w_{t-1}, f_{t-1}(w_{t-1})\}$. The player uses an algorithm $\mathcal{A}$ to sequentially pick w_t based on the available information I_t, i.e., $\mathcal{A}(I_t) = w_t$. We can overload the definition of regret to represent the regret of an algorithm:

$$\mathcal{R}_T(\{f_t\}_{t=1}^T, \mathcal{A}, w) = \sum_{t=1}^{T} f_t(w_t) - f_t(w) \quad \text{such that} \quad \mathcal{A}(\mathcal{I}_t) = w_t$$

Assume that the functions $f_1, \ldots, f_T$ belong to a function class $\mathcal{F}$. An algorithm $\mathcal{A}$ is considered *no-regret* for the sets $\mathcal{F}, \mathcal{D}$ if it guarantees that the regret incurred relative to any point in $\mathcal{D}$ grows at a sub-linear rate in T for any sequence of functions from $\mathcal{F}$. Mathematically,

$$\lim_{T\to\infty} \frac{\mathcal{R}_T(\{f_t\}_{t=1}^T, \mathcal{A}, w)}{T} = 0 \quad \text{for all} \quad f_1, \ldots, f_T \in \mathcal{F}, w \in \mathcal{D}$$

This implies that the player's average regret per round approaches zero for any sequence of functions from $\mathcal{F}$.

For a particular online optimization problem characterized by $\mathcal{F}$ and $\mathcal{D}$, one can analyze a candidate algorithm $\mathcal{A}$ and show that it has the no-regret property. Further, one can also analyze the *worst-case* asymptotic growth rate of regret in the Big-O notation, providing a concrete regret upper-bound for $\mathcal{A}$. Finally, one can show a lower bound, establishing the least regret any algorithm must incur for the problem. If the algorithm's regret upper-bound matches the problem's regret lower-bound, then the algorithm has *minimax optimal* regret for the problem. Indeed, much of the early literature on online optimization focused on designing and analyzing minimax optimal algorithms.

However, such worst-case regret bounds apply uniformly for all sequences of functions in $\mathcal{F}$.

The bound could be overly conservative for the particular sequence of realized functions $f_1, \ldots, f_t$. This has led to the study of *data-dependent* regret bounds and *adaptive* algorithms. These algorithms hope to bound the regret as an explicit function of $f_1, \ldots, f_t$. If the realized sequence of functions is generated adversarially, these regret bounds should imply the worst-case bounds. On the other hand, if the sequence is benign, the bound should be lower than the worst-case bounds. This flexibility makes adaptive algorithms desirable for practical applications where the data may not be adversarially generated but is challenging to model accurately. This dissertation studies data-dependent regret bounds and adaptive algorithms for two online learning problems, namely the *Adversarial Multi-Armed Bandits* [11] in Chapter 3, the *Online Portfolio Selection* [1] problem in Chapter 4 and Chapter 5.

Consider an extension of the online learning framework, where at the beginning of round t, the player is given a *predicted function* m_t, which is supposed to be a prediction of the actual cost function f_t which is yet to be seen. The player then selects w_t using the augmented information set $\mathcal{I}_t \cup \{m_t\}$. A *greedy* strategy for the player is to pick $w_t \in \arg\min_{w \in \mathcal{D}} m_t(w)$. If the predicted function m_t is a close approximation of the realized function f_t, then the greedy strategy performs very well. However, the player's performance would degrade if the predicted function is inaccurate. The greedy strategy may not even have the no-regret property. On the other hand, the player could employ a no-regret strategy that uses the information set $\mathcal{I}_t$, completely ignoring the prediction m_t. While the no-regret strategy comes with a guaranteed performance bound, it is unable to take advantage of m_t if it is indeed an accurate prediction of f_t. A recent active area of research called *online learning with predictions* (OLP) [12], studies online learning algorithms that are augmented with predictions. The particular details of how the predicted function m_t is created are not a subject of study in the framework. Instead, it studies algorithms that can use predictions, if available, to improve regret bounds. The OLP framework provides another technique besides data-dependent regret bounds for going beyond the worst-case regret bounds.

The goal of the OLP framework is to design algorithms that have two principal properties. First, they should be *consistent*, i.e., if the predictions are accurate, the algorithm should be able

to take advantage of this and have better than worst-case performance. Second, they should be *robust*, i.e., if the predictions are entirely arbitrary, the algorithm's performance should be similar to the worst-case algorithm that does not use any predictions. This dissertation presents algorithms for the online portfolio selection problem with predictions that exhibit the consistency-robustness properties in Chapter 5.

The algorithms developed in this dissertation are based on the Follow-The-Regulazied-Leader (FTRL) technique from OCO. In Chapter 2, we present an overview of FTRL along with a general regret inequality that uses the complete function f_t, incorporates arbitrary convex predictions m_t, and employs a changing learning rate. All the regret bounds in subsequent chapters can be obtained by applying a suitable version of this regret inequality. We use the AdaFTRL [13] technique for optimally tuning learning rates in various problems. Our analysis relies on convex analysis techniques and extensively uses the *Bregman divergence*. We present a novel technique for obtaining local-norm lower bounds for Bregman divergences that plays a crucial role in obtaining our regret bounds. Our general regret inequality and the Bregman divergence lower-bound could be of independent technical interest.

Chapter 2: Online Convex Optimization with Convex Predictions

2.1 Introduction

The Online Convex Optimization framework (OCO) was first defined by Zinkevich [10]. It provides a powerful framework for the design and analysis of regret minimizing algorithms. In the last two decades, there have been many developments in this area and it continues to be an active area of research within the the machine learning, operations research and statistics communities. It has also seen widespread adoption by practitioners. Algorithms that originate from OCO, like AdaGrad [14] and Adam [15] are widely used as optimizers for training deep neural networks. The monographs of Hazan [6], Shalev-Shwartz [9], and Orabona [8] provide a comprehensive overview of OCO. For the sake of completeness, we re-phrase the OCO interaction protocol from Chapter 1.

A player interacts with an environment for T rounds. In each round, the player selects an action from a convex set, $w_t \in \mathcal{D} \subseteq \mathbb{R}^n$. The environment picks a convex function $f_t : \mathcal{D} \to \mathbb{R}$. The player incurs a scalar cost $f_t(w_t)$ and observes the function f_t. The player's objective is to minimize the total cost of interaction over the T rounds $\sum_{t=1}^{T} f_t(w_t)$. The *static-regret* (regret for short) of the player compared to the cost of fixed point $w \in \mathcal{D}$ is $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$. The action w_t is selected using an algorithm $\mathcal{A}$, that takes as input the current information set $\mathcal{I}_t = \{w_1, f_1, \ldots, w_{t-1}, f_{t-1}\}$ and outputs the action, i.e. $w_t = \mathcal{A}(I_t)$. The interaction protocol is summarized below:

Online Convex Optimization - Interaction Protocol:

Initial information set $\mathcal{I}_1 = \{\}$

for $t = 1$ *to* T **do**

- Player picks $w_t = \mathcal{A}(\mathcal{I}_t)$
- Environment picks f_t
- Player incurs cost $f_t(w_t)$
- Update information set $\mathcal{I}_{t+1} = I_t \cup \{w_t, f_t\}$

A straightforward strategy for the player is to select w_t using *Follow The Leader* (FTL). FTL can be succinctly expressed as:

$$w_t \in \arg\min_{w \in \mathcal{D}} \sum_{s=1}^{t-1} f_s(w) \tag{FTL}$$

Unfortunately, FTL can have $O(T)$ regret even with linear functions [8, Example 2.10]. This occurs because FTL's iterates can be forced into alternating between opposite corners of $\mathcal{D}$ in every iteration, making it *"unstable"*. Nevertheless, FTL has $O(\log T)$ regret when the functions are strongly convex [8, Corollary 7.24]. Even for linear functions, FTL's regret is $O(\log T)$ if the decision set's boundary exhibits sufficient curvature [16, 17].

Incorporating regularization into FTL is a common approach to *"stabilize"* the iterations, resulting in a widely studied algorithm in the OCO literature called Follow The Regularized Leader (FTRL). Another popular algorithm for OCO is Online Mirror Descent (OMD), which stabilizes the iterates by ensuring consecutive iterates remain close to each other. Several fascinating connections and equivalences exist between FTRL and OMD, as discussed in [8]. Various well-known iterative algorithms in machine learning, such as Online Gradient Descent [10], AdaGrad [14], Exponentiated Gradient[18], and Online Newton Step [3], can be formulated using one of these two algorithms. See [8] for a detailed history of FTRL. In this thesis, we focus on the FTRL algorithm and its variants. In its simplest form, it can be stated as:

$$w_t \in \arg\min_{w \in \mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + \frac{F(w)}{\eta_{t-1}} \tag{FTRL}$$

Here, $F(w)$ is the regularization function and η_{t-1} is a time varying learning-rate parameter that needs to be tuned. While there are several techniques for picking η_{t-1}, we focus on the AdaFTRL technique prescribed by Orabona and Pál [13] for obtaining data-dependent regret bounds.

The above FTRL requires minimizing the sum of t convex functions in round t. In general, this could potentially require $O(t)$ computation per round, becoming increasingly costly as the number of rounds increases. One can construct *surrogate* convex functions that are linear or quadratic and

run FTRL on them to mitigate this computational issue. Assume we have a surrogate function $\tilde{f}_t$ such that $f_t(w_t) = \tilde{f}_t(w_t)$ and $f_t(w) \geq \tilde{f}_t(w)$ for all $w \in \mathcal{D}$, then we have:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w)$$

Thus, if we can bound the regret of the surrogates $\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w)$ using FTRL on $\tilde{f}_t$, we obtain a bound on the regret $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$. Running an FTRL on $\tilde{f}_t$ instead of f_t not only delivers computational benefits, but may also aid in obtaining tighter regret bounds. We will see this happen for the online portfolio slection problem in Chapter 4.

In the online learning with predictions (OLP) framework, the player is given a predicted function $m_t(w)$ before picking w_t. The interaction protocol for OLP is below:

Online Convex Optimization with Predictions - Interaction Protocol:

Initial information set $\mathcal{I}_1 = \{\}$

for $t = 1$ *to* T **do**

- Receive prediction m_t
- Player picks $w_t = \mathcal{A}(\mathcal{I}_t \cup \{m_t\})$
- Environment picks f_t
- Player incurs cost $f_t(w_t)$
- Update information set $\mathcal{I}_{t+1} = I_t \cup \{w_t, f_t\}$

For OLP, we employ the Optimistic FTRL [12] algorithm:

$$w_t \in \arg\min_{w \in \mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + m_t(w) + \frac{F(w)}{\eta_{t-1}} \tag{OFTRL}$$

The OLP framework, while formally introduced in [12, 19], had previously appeared in various forms [20, 21]. This framework has been instrumental in demonstrating several intriguing results, such as adaptive regret bounds in online learning [22], adaptive regret bounds in adversarial bandits [23], and accelerated rates of convergence for two-player games [24], to name a few. A related area of research, called *Algorithms with Predictions* [25] studies how predictions could be used to improve the performance of online algorithms, where the performance benchmark is competitive

ratio. While the problems studied in this area are different from OLP, they share the common goal of going beyond wort-case performance with the help of predictions.

2.2 Regret Inequality for Optimistic FTRL with Convex Predictions

In prior works such as Rakhlin and Sridharan [12], the regret inequality is obtained for linear costs and linear predictions (See Luo [26] for a simple proof). We extend their result and obtain a general regret inequality for Optimistic FTRL with convex cost functions f_t and convex predictions m_t. Our result is stated in terms of *Bregmen Divergences* and *Mixed-Bregmans*.

Definition 2.1 (Bregman Divergence)**.** *The Bregman Divergence of function F is:*

$$B_F(x\|y) = F(x) - F(y) - \nabla F(y)^\top (x - y)$$

Definition 2.2 (Mixed Bregman)**.** *For $\alpha, \beta > 0$ the (α, β)-Mixed Bregman of function F is:*

$$B_F^{\alpha,\beta}(x\|y) = \frac{F(x)}{\alpha} - \frac{F(y)}{\beta} - \frac{\nabla F(y)}{\beta}^\top (x - y)$$

The Mixed Bregman is not a divergence as $\mathrm{B}_F^{\alpha,\beta}(x\|x)$ may not be zero. However, we do have the relation $\alpha \mathrm{B}_F^{\alpha,\alpha}(x\|y) = \mathrm{B}_F(x\|y)$.

Let the iterates of Optimistic FTRL be w_t:

$$w_t \in \arg\min_{w \in \mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + m_t(w) + \frac{F(w)}{\eta_{t-1}}$$

Let the iterates of FTRL be w'_t:

$$w'_t \in \arg\min_{w \in \mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + \frac{F(w)}{\eta_{t-1}}$$

We will use the shorthand $g_t = \sum_{s=1}^{t} f_s$. The following theorem bounds the regret of Optimistic FTRL in terms of the iterates w_t and w'_t. The proof appears in Appendix A

Theorem 2.3. *For any $w \in \mathcal{D}$, any sequence of convex cost functions $f_1, \ldots, f_T$, convex hint functions $m_1, \ldots, m_T$, convex regularizer F and parameters $\eta_0, \ldots, \eta_T$ such that $w_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + m_t(w) + \frac{F(w)}{\eta_{t-1}}$ and $w'_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + \frac{F(w)}{\eta_{t-1}}$. Let $g_t = \sum_{s=1}^{t} f_s$. The iterates of Optimistic FRTL $w_1, \ldots, w_T$ satisfies the regret inequality $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:*

$$\leq B_F^{\eta_T,\eta_0}(w\|w'_1) + \sum_{t=1}^{T}\Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - B_{g_t}(w'_{t+1}\|w_t) - B_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t) \\ - B_{g_{t-1}}(w_t\|w'_t) - B_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t)\Big]$$

Further, if F is such that $min_{w\in\mathcal{D}} F(w) = 0$ and the sequence $\eta_0, \ldots, \eta_T$ is non-increasing, then the above bound simplifies to $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:

$$\leq \frac{F(w)}{\eta_T} + \sum_{t=1}^{T}\Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - B_{g_t}(w'_{t+1}\|w_t) - \frac{B_F(w'_{t+1}\|w_t)}{\eta_{t-1}} \\ - B_{g_{t-1}}(w_t\|w'_t) - \frac{B_F(w_t\|w'_t)}{\eta_{t-1}}\Big]$$

In most applications, including the ones in this thesis, we typically ignore the last two terms in the summation and use the inequality $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:

$$\leq \frac{F(w)}{\eta_T} + \sum_{t=1}^{T}\Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1}\|w_t) - \frac{\mathrm{B}_F(w'_{t+1}\|w_t)}{\eta_{t-1}}\Big] \tag{2.1}$$

However, these two terms do play a role in certain applications, like in showing convergence in general convex games Farina *et al.* [27], data-dependent regret bounds in Multi-Armed Bandits Wei and Luo [23] and gradual-variation bounds in online learning Chiang *et al.* [21]. In Orabona [8], a similar regret bound for Optimistic FTRL is obtained, which when translated into our notation would imply $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:

$$\leq \frac{F(w)}{\eta_T} + \sum_{t=1}^{T}\Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w_{t+1}) - \mathrm{B}_{g_t}(w_{t+1}\|w_t) - \frac{\mathrm{B}_F(w_{t+1}\|w_t)}{\eta_{t-1}}\Big] \tag{2.2}$$

In our bound, we bound the regret of Optimistic FTRL in terms of iterates of both Optimistic FTRL w_t and FTRL w'_t. Whereas in Orabona [8], only the iterates of Optimistic FTRL appear. In our analysis of Optimistic FTRL, we separate the hint $m_t(w)$ from the regularizer term $F(w)/\eta_{t-1}$. On the other hand, in Orabona [8], the sum $m_t(w)+F(w)/\eta_{t-1}$ is treated as a composite regularizer and analyzed using their FTRL bound. Our general regret bound in Theorem 2.3 is novel as it could be useful in applications that require the last two terms, like [27, 23]. These two terms cannot be obtained via the analysis in [8].

In the case where we have no hints, i.e., $m_t = 0$, Equation (2.1) and Equation (2.2) become equivalent to the well known FTRL regret inequality. The iterates of FTRL are given by:

$$w_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + \frac{F(w)}{\eta_{t-1}}$$

The regret of FTRL is stated in the following Corollary.

Corollary 2.4. *For any $w \in \mathcal{D}$, any sequence of convex cost functions $f_1, \ldots, f_T$ and parameters $\eta_0, \ldots, \eta_T$ such that $w_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + \frac{F(w)}{\eta_{t-1}}$. Assume F is such that $min_{w\in\mathcal{D}} F(w) = 0$ Let $g_t = \sum_{s=1}^{t} f_s$. The iterates of FRTL $w_1, \ldots, w_T$ satisfies the regret inequality:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \le \frac{F(w)}{\eta_T} + \sum_{t=1}^{T} \left[\nabla f_t(w_t)^\top (w_t - w_{t+1}) - B_{g_t}(w_{t+1}\|w_t) - \frac{B_F(w_{t+1}\|w_t)}{\eta_{t-1}}\right]$$

Proof. When $m_t = 0$, the iterates of Optimistic FTRL w_t and FTRL w'_t coincide. So, the last two terms in the result of Theorem 2.3 vanish, i.e., $\mathrm{B}_{g_{t-1}}(w_t\|w'_t) = 0$ and $\mathrm{B}_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t) = 0$. ■

In some applications, besides the regularizer F whose strength is regulated through η_t, we may need to add an extra constant regularizer G to the optimization. The update equation here is:

$$w_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + G(w) + \frac{F(w)}{\eta_{t-1}}$$

Note that this is different from the optimistic FTRL update where the hint m_t may change over

time. In the above update, G is treated as a regularizer and not as a hint. The regret inequality for this update is:

Corollary 2.5. *For any $w \in \mathcal{D}$, any sequence of convex cost functions $f_1, \ldots, f_T$ and parameters $\eta_0, \ldots, \eta_T$ such that $w_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + G(w) + \frac{F(w)}{\eta_{t-1}}$. Assume G, F are such that $min_{w\in\mathcal{D}} G(w) = 0$ and $min_{w\in\mathcal{D}} F(w) = 0$. Let $g_t = \sum_{s=1}^{t} f_s$. The iterates $w_1, \ldots, w_T$ satisfies the regret inequality $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:*

$$\leq G(w) + \frac{F(w)}{\eta_T} + \sum_{t=1}^{T} \left[\nabla f_t(w_t)^\top (w_t - w_{t+1}) - B_{g_t}(w_{t+1} \| w_t) - B_G(w_{t+1} \| w_t) - \frac{B_F(w_{t+1} \| w_t)}{\eta_{t-1}} \right]$$

Proof. We can apply Corollary 2.4 starting from time $t = 0$. We use $f_0(w) = G(w)$, and $\eta_{-1} > 0$ in Corollary 2.4. This gives the regret inequality:

$$\sum_{t=0}^{T} f_t(w_t) - f_t(w) \leq \frac{F(w)}{\eta_T} + \sum_{t=0}^{T} \left[\nabla f_t(w_t)^\top (w_t - w_{t+1}) - \mathrm{B}_{g_t}(w_{t+1} \| w_t) - \mathrm{B}_G(w_{t+1} \| w_t) \frac{\mathrm{B}_F(w_{t+1} \| w_t)}{\eta_{t-1}} \right]$$

We can write the left hand side of the above inequality as:

$$\sum_{t=0}^{T} f_t(w_t) - f_t(w) = G(w_0) - G(w) + \sum_{t=1}^{T} f_t(w_t) - f_t(w)$$

On the right hand side, we simply the term inside the sum when $t = 0$ as:

$$\begin{aligned} \nabla G(w_0)^\top (w_0 - w_1) - \mathrm{B}_G(w_1 \| w_0) - \frac{\mathrm{B}_F(w_1 \| w_0)}{\eta_{-1}} &\leq \nabla G(w_0)^\top (w_0 - w_1) - \mathrm{B}_G(w_1 \| w_0) \\ &= G(w_0) - G(w_1) \leq G(w_0) \end{aligned}$$

For $t = 1 \ldots T$, the term inside the sum is:

$$\nabla f_t(w_t)^\top (w_t - w_{t+1}) - \mathrm{B}_{g_t}(w_{t+1} \| w_t) - \mathrm{B}_G(w_{t+1} \| w_t) - \frac{\mathrm{B}_F(w_{t+1} \| w_t)}{\eta_{t-1}}$$

Putting the two sides together and simplifying, we get the stated result. ∎

2.3 Tuning η_t using the AdaFRTL technique

In Theorem 2.3, Corollary 2.4 and Corollary 2.5, the inequalities contain the following common form for suitable of A, b_t and C_t.

$$\frac{A}{\eta_T} + \sum_{t=1}^{T} b_t^\top (w_t - w_{t+1}) - \mathrm{B}_{C_t(\eta_{t-1})}(w_{t+1} \| w_t)$$

The AdaFTRL strategy picks a specific sequence of parameters η_{t-1} based on the history $\mathcal{I}_t$. This strategy was analyzed in Orabona and Pál [13] and a simpler analysis was given by Koolen [28]. They give a simple algorithmic technique for tuning η_{t-1}. Our analysis is adapted from Hadiji and Stoltz [29]. We consider time varying parameters of the form:

$$\eta_t = \frac{\alpha}{\beta + \sum_{s=1}^{t} M_s(\eta_{s-1})}$$

Where $\alpha, \beta > 0$ are constants and $M_t(\eta)$ is the optimal value of the following optimization.

$$M_t(\eta) = \sup_{w \in \mathcal{D}} b_t^\top (w_t - w) - \mathrm{B}_{C_t(\eta)}(w \| w_t)$$

Thus, we have the sum:

$$\frac{A}{\eta_T} + \sum_{t=1} b_t^\top (w_t - w_{t+1}) - \mathrm{B}_{C_t(\eta_{t-1})}(w_{t+1} \| w_t) \leq \frac{A}{\eta_T} + \sum_{t=1}^{T} M_t(\eta_{t-1})$$

We bound the above sum using the following lemma:

Lemma 2.6. *Let* $\eta_t = \frac{\alpha}{\beta + \sum_{s=1}^{t} M_s(\eta_{s-1})}$. *If* $0 \leq M_t(\eta_{t-1}) \leq L$ *for all* $t = 1, \ldots T$ *and* $\frac{M_t(\eta_{t-1})}{\eta_{t-1}} \leq g_t$, *then we have the upper bound:*

$$\frac{A}{\eta_T} + \sum_{t=1}^{T} M_t(\eta_{t-1}) \leq A\left(\frac{\beta}{\alpha} + \frac{L}{\alpha}\right) + L + \sqrt{2\sum_{t=1}^{T} g_t}\left(\frac{A}{\sqrt{\alpha}} + \sqrt{\alpha}\right)$$

Proof. Substituting for η_T , we have:

$$\frac{A}{\eta_T} + \sum_{t=1}^{T} M_t(\eta_{t-1}) = \frac{A\beta}{\alpha} + \left(\frac{A}{\alpha} + 1\right) \sum_{t=1}^{T} M_t(\eta_{t-1})$$

Consider $\left(\sum_{t=1}^{T} M_t(\eta_{t-1})\right)^2$

$$\begin{aligned}
\left(\sum_{t=1}^{T} M_t(\eta_{t-1})\right)^2 &= \sum_{t=1}^{T} M_t(\eta_{t-1})^2 + 2 \sum_{t=1}^{T} M_t(\eta_{t-1}) \sum_{s=1}^{t-1} M_s(\eta_{s-1}) \\
&= \sum_{t=1}^{T} M_t(\eta_{t-1})^2 + 2 \sum_{t=1}^{T} M_t(\eta_{t-1}) \left(\frac{\alpha}{\eta_{t-1}} - \beta\right) \\
&\leq \sum_{t=1}^{T} M_t(\eta_{t-1})^2 + 2\alpha \sum_{t=1}^{T} \frac{M_t(\eta_{t-1})}{\eta_{t-1}} \\
&\leq L \sum_{t=1}^{T} M_t(\eta_{t-1}) + 2\alpha \sum_{t=1}^{T} g_t
\end{aligned}$$

Using the fact that $x^2 \leq a + bx$ implies that $x \leq \sqrt{a} + b$ for all $a, b, x \geq 0$, we have:

$$\sum_{t=1}^{T} M_t(\eta_{t-1}) \leq \sqrt{2\alpha \sum_{t=1}^{T} g_t} + L$$

Thus, we get:

$$\begin{aligned}
&\frac{A\beta}{\alpha} + \left(\frac{A}{\alpha} + 1\right) \sum_{t=1}^{T} M_t(\eta_{t-1}) \leq \frac{A\beta}{\alpha} + \left(\frac{A}{\alpha} + 1\right) \left(\sqrt{2\alpha \sum_{t=1}^{T} g_t} + L\right) \\
&= A\left(\frac{\beta}{\alpha} + \frac{L}{\alpha}\right) + L + \sqrt{2 \sum_{t=1}^{T} g_t} \left(\frac{A}{\sqrt{\alpha}} + \sqrt{\alpha}\right)
\end{aligned}$$

■

The constants α and β are tuning based on A and L in order to obtain a concrete bound.

2.4 Logarithmic-Barrier Regularizer

The final piece is the regularizer. For the applications in this thesis, the regularizer we use is the *Logarithmic-Barrier*. Let Δ_n be the probability simplex $\{x \in \mathbb{R}^n : \sum_{i=1}^n x(i) = 1, x(i) \geq 0, i \in [n]\}$. The log-barrier regularizer defined on Δ_n is given by the function:

$$F(x) = -n\log(n) - \sum_{i=1}^{n} \log(x(i))$$

We explore a few important properties of the log-barrier here.

Definition 2.7 (Legendre function)**.** *A continuous function $F : \mathcal{D} \to \mathbb{R}$ is Legendre if F is strictly convex, continuously differentiable on Interior($\mathcal{D}$) and* $\lim_{x \to \mathcal{D}/Interior(\mathcal{D})} \|\nabla F(x)\| = +\infty$.

It is easy to verify that the log-barrier is a Legendre function on the domain $\mathbb{R}^n_+$.

A crucial step in the analysis of FTRL involves bounding the so-called stability term $\Psi_x(l)$, which is defined as:

$$\Psi_x(l) = \sup_{y \in \Delta_n} l^\top (x - y) - \mathrm{B}_F(y\|x)$$

Let $y^\star$ be the point in Δ_n achieving the supremum in the definition of $\Psi_x(l)$. As F is Legendre, the supremum is always attained at a unique $y^\star$ in Δ_n.

Let $H(x)$ be a positive definite matrix for every $x \in \Delta_n$. Define the norm $\|z\|^2_{H(x)} = z^\top H(x) z$. We call such norms as *local norms*. Let ω is a non-negative convex function. Suppose a lower bound of the following form holds for all $x, y \in \Delta_n$:

$$\mathrm{B}_F(y\|x) \geq \omega(\|x - y\|_{H(x)})$$

Then, we obtain the following upper-bound for $\Psi_x(l)$:

$$\Psi_x(l) = l^\top (x - y^\star) - \mathrm{B}_F(y^\star\|x) \leq \|l\|_{H(x)^{-1}} \|x - y^\star\|_{H(x)} - \omega(\|x - y^\star\|_{H(x)}) \leq \omega^\star(\|l\|_{H(x)^{-1}})$$

Here $\omega^\star$ is the Fenchel-dual of ω, given by $\omega^\star(t) = \sup_s(st - \omega(s))$.

Using the theory of *self-concordant functions* [30], it is possible to obtain one such bound for $\Psi_x(l)$.

Definition 2.8 (Self-Concordant Function)**.** *A continuous function $F : \mathcal{D} \to \mathbb{R}$ is M self-concordant if F is a Legendre function on $\mathcal{D}$ and satisfies:*

$$|\nabla^3 F(x)[u,u,u]| \leq 2M(\nabla^2 F(x)[u,u])^{3/2} \quad \forall x \in \mathcal{D}, u \in \mathbb{R}^n$$

It is easy to verify that the log-barrier satisfies the self-concordance condition with $M = 1$.

Nesterov [30] obtains the following lower bound for $\mathrm{B}_F(y\|x)$.

Lemma 2.9 (Theorem 5.1.8, [30])**.** *For any $x, y \in \mathbb{R}^n_+$ and $F(x) = -n\log(n) - \sum_{i=1}^n \log(x(i))$, we have:*

$$B_F(y\|x) \geq \omega(\|x-y\|_{\nabla^2 F(x)})$$

Here $\omega(t) = t - \log(1+t)$.

Using Lemma 2.9, we have the following theorem bounding $\Psi_x(l)$.

Lemma 2.10. *Let $F(x) = -n\log(n) - \sum_{i=1}^n \log(x(i))$. For all $x, y \in \Delta_n$ and $l \in \mathbb{R}^n$ such that $\|l\|_{\nabla^2 F(x)^{-1}} \leq 1$, we have the upper-bound:*

$$l^\top(x-y) - B_F(y\|x) \leq \omega^\star(\|l\|_{\nabla^2 F(x)^{-1}})$$

where $\omega^\star(t) = -t - \log(1-t)$. Further, if we have $\|l\|_{\nabla^2 F(x)^{-1}} \leq \frac{1}{2}$, then we have:

$$l^\top(x-y) - B_F(y\|x) \leq \|l\|^2_{\nabla^2 F(x)^{-1}} = \sum_{i=1}^n x(i)^2 l(i)^2$$

Proof. Using Holder's inequality and Lemma 2.9, we have:

$$l^\top(x-y) - \mathrm{B}_F(y\|x) \leq \|l\|_{\nabla^2 F(x)^{-1}}\|x-y\|_{\nabla^2 F(x)} - \omega(\|x-y\|_{\nabla^2 F(x)}) \leq \omega^\star(\|l\|_{\nabla^2 F(x)^{-1}})$$

Here $\omega^\star(t) = -t - \log(1-t)$ is the Fenchel-dual of ω. Since the domain of $\omega^\star$ is $(-\infty, 1)$, the above inequality holds when $\|l\|_{\nabla^2 F(x)^{-1}} < 1$. Observe that when $t \in [0, 1/2]$, $\omega^\star(t) \leq t^2$. So, when $\|l\|_{\nabla^2 F(x)^{-1}} \leq 1/2$, we have the bound:

$$l^\top (x - y) - \mathrm{B}_F(y\|x) \leq \|l\|^2_{\nabla^2 F(x)^{-1}} = \sum_{i=1}^{n} x(i)^2 l(i)^2$$

■

In applications within online learning, Lemma 2.10 is typically used when $\|l\|_{\nabla^2 F(x)^{-1}} \leq \frac{1}{2}$ holds. We use this result in Chapter 4 and Chapter 5 for the Online Portfolio Selection problem. In the next subsection, we show a new bound that holds for any $l \in \mathbb{R}^n$. Our new bound plays an important role in Chapter 3 on the Adversarial Multi-Armed Bandit problem.

2.5 New Local-norm Lower bounds for Bregman Divergences

We use *Potential Functions* to construct our lower bounds. These were introduced in [31, 32, 33] for analyzing the regret of online learning algorithms.

Definition 2.11 (Potential Function)**.** *A function $\psi : (-\infty, a) \to (0, +\infty)$ for some $a \in \mathbb{R} \cup \{+\infty\}$ is called a Potential if it is convex, strictly increasing, continuously differentiable and satisfies:*

$$\lim_{x \to -\infty} \psi(x) = 0 \quad \textit{and} \quad \lim_{x \to a} \psi(x) = +\infty$$

For instance, $\exp(x)$ is a potential with $a = \infty$ and $-1/x$ is a potential with $a = 0$. Associated with a potential ψ, we define a function f_ψ as the indefinite integral $f_\psi(z) = \int \psi^{-1}(z)dz + C$. Since the domain of ψ^{-1} is $(0, \infty)$, the domain of f_ψ is also $(0, \infty)$. For instance, if $\psi(x) = -1/x$ on the domain $(-\infty, 0)$, the associated function is $f_\psi(x) = -\log(x) + C$.

Observe that $f'_\psi(z) = \psi^{-1}(z)$ and $f''_\psi(z) = \left[\psi'(\psi^{-1}(z))\right]^{-1}$. Since ψ is strictly convex and increasing, $\psi' > 0$ and thus $f''_\psi > 0$, making f_ψ strictly convex. Moreover, $\lim_{z \to 0} \mid f'_\psi(z) \mid = \lim_{z \to 0} \mid \psi^{-1}(z) \mid = +\infty$. Define the function $F_\psi : \mathbb{R}^n \to \mathbb{R}$ as $F_\psi(x) = \sum_{i=1}^n [f_\psi(x(i)) - f_\psi(1/n)]$.

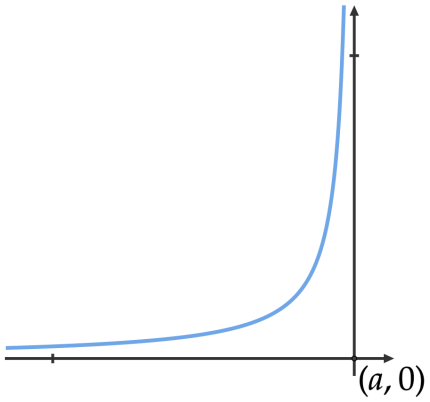

Figure 2.1: Potential Function

Given a potential $\psi : (-\infty, a) \to (0, +\infty)$ and its associated function f_ψ, the Fenchel dual of f_ψ is $f_\psi^\star : (-\infty, a) \to \mathbb{R}$ defined as $f_\psi^\star(u) = \sup_{z>0}(zu - f_\psi(z))$. The supremum is achieved at $z = {f'_\psi}^{-1}(u) = \psi(u)$. So we have that $f_\psi^\star(u) = u\psi(u) - f_\psi(\psi(u))$. This implies $f_\psi^{\star\prime}(u) = \psi(u)$ and $f_\psi^{\star\prime\prime}(u) = \psi'(u)$. Further, using integration by parts on $\int \psi(u)du$ and substituting $\psi(u) = s$:

$$\int \psi(u)du = u\psi(u) - \int u\psi'(u)du = u\psi(u) - \int \psi^{-1}(s)ds = u\psi(u) - f_\psi(\psi(u)) + C = f_\psi^\star(u) + C$$

Thus $f_\psi^\star(u) = \int \psi(u)du - C$. Here C is the same constant of integration picked when defining $f_\psi(z) = \int \psi^{-1}(z)dz + C$. We have the following property:

Lemma 2.12. *Let x, y be such that $x = \psi(u)$ and $y = \psi(v)$. Then $B_{f_\psi}(y\|x) = B_{f_\psi^\star}(u\|v)$*

Proof. Use the fact that $f_\psi^\star(u) = u\psi(u) - f(\psi(u))$.

$$\begin{aligned}
\mathrm{B}_{f_\psi}(y\|x) = \mathrm{B}_{f_\psi}(\psi(v)\|\psi(u)) &= f_\psi(\psi(v)) - f_\psi(\psi(u)) - f'_\psi(\psi(u))(\psi(v) - \psi(u)) \\
&= v\psi(v) - f_\psi^\star(v) - (u\psi(u) - f_\psi^\star(u)) - u(\psi(v) - \psi(u)) \\
&= f_\psi^\star(u) - f_\psi^\star(v) - f^{\star\prime}{}_\psi(v)(u - v) = \mathrm{B}_{f_\psi^\star}(u\|v)
\end{aligned}$$

■

The following lemma obtains the local-norm lower-bounds.

Lemma 2.13. *Let ψ be a potential and $x \in \mathbb{R}_+$ such that $x = \psi(u)$ for some u. Let ϕ be a non-negative function such that $\psi(u + \phi(u))$ exists. Define the function $h(z) = \frac{\psi(z+\phi(z))-\psi(z)}{\phi(z)}$. For all $0 < y \leq \psi(u + \phi(u))$ we have the lower bound:*

$$B_{f_\psi}(y\|x) \geq \frac{1}{2}\frac{(x-y)^2}{h(\psi^{-1}(x))}$$

Proof. Let v be such that $y = \psi(v)$. Using Lemma 2.12, we have $\mathrm{B}_{f_\psi}(y\|x) = \mathrm{B}_{f_\psi^\star}(u\|v)$. Using the fact that $f_\psi^\star(u) = \int \psi(u)du - C$, we have:

$$\mathrm{B}_{f_\psi^\star}(u\|v) = f_\psi^\star(u) - f_\psi^\star(v) - f_\psi^{\star\prime}(v)(u-v) = \int_v^u \psi(s) - y(u-v)$$

We can visualize $\mathrm{B}_{f_\psi^\star}(u\|v)$ using the potential function. When $v \leq u$, it is the area with green borders in Figure 2.2 and when $u \leq v$, it is the area with green borders in Figure 2.3. Consider

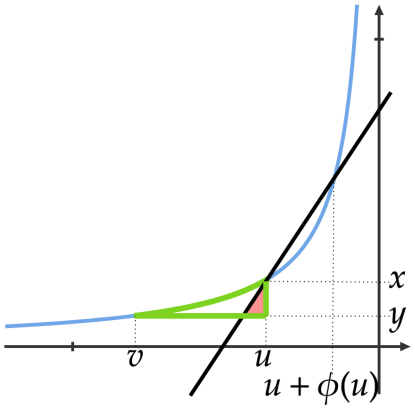

Figure 2.2: $v \leq u$

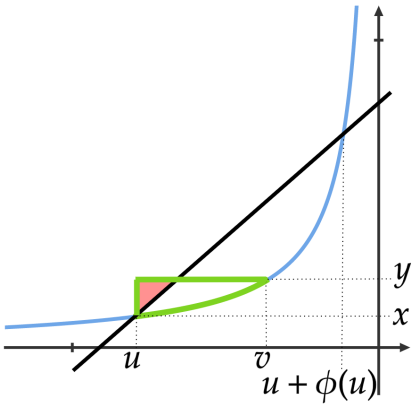

Figure 2.3: $u \leq v \leq u + \phi(u)$

the line passing through (u, x) and $(u + \phi(u), \psi(u + \phi(u)))$. Its slope is $h(u) \geq \psi'(u) > 0$. In both cases, the height of the red triangle is $|x - y|$ and its base is $\frac{|x-y|}{h(u)}$. So, the area of the red triangle will be $\frac{1}{2}\frac{(x-y)^2}{h(u)}$. Since the triangle is always smaller than $\mathrm{B}_{f_\psi^\star}(u\|v)$, we have the lower bound $\mathrm{B}_{f_\psi}(y\|x) \geq \frac{1}{2}\frac{(x-y)^2}{h(\psi^{-1}(x))}$. ■

In the context of online learning, local-norm lower-bounds have been studied before, see for example Orabona [8]. However, these rely upon Taylor's theorem to show that $\mathrm{B}_{f_\psi}(y\|x) = \frac{1}{2}(x -$

$y)^2 f''_\psi(z)$ for some $z \in [x, y]$. Then, they use further conditions on x, y to argue that $c f''_\psi(x) \leq f''_\psi(z)$ for some positive constant c and thus arrive at $\mathrm{B}_{f_\psi}(y\|x) \geq \frac{c}{2}(x-y)^2 f''_\psi(x)$. We generalize this argument in Lemma 2.13, through which we are able to generate a more rich class of lower-bounds. We illustrate the use of our theorem via an example:

Corollary 2.14. *Let $\psi(u) = -1/u$ in the domain $(-\infty, 0)$. So $f_\psi(x) = -\log(x) + C$. For $x, y \in (0, 1]$, we have the lower-bound*

$$B_{f_\psi}(y\|x) = \frac{y}{x} - 1 - \ln\left(\frac{y}{x}\right) \geq \frac{1}{2}\frac{(x-y)^2}{x}$$

Proof. For any $x \in (0, 1]$, let $u \in (-\infty, -1]$ be such that $\psi(u) = x$. Let $\phi(u) = -1 - u$. Clearly, $\phi(u) \geq 0$ and $\psi(u + \phi(u)) = \psi(-1) = 1$. We have

$$h(u) = \frac{\psi(u + \phi(u)) - \psi(u)}{\phi(u)} = \frac{1 + \frac{1}{u}}{-1 - u} = \frac{-1}{u} = \psi(u) = x$$

Applying Lemma 2.13, we have the lower-bound for all $0 < y \leq 1$:

$$\mathrm{B}_{f_\psi}(y\|x) = \frac{y}{x} - 1 - \ln\left(\frac{y}{x}\right) \geq \frac{1}{2}\frac{(x-y)^2}{h(\psi^{-1}(x))} = \frac{1}{2}\frac{(x-y)^2}{x}$$

■

Using Corollary 2.14, we can construct the following new upper-bound:

Lemma 2.15. *Let $F(x) = -n\log(n) - \sum_{i=1}^n \log(x(i))$. For all $x, y \in \Delta_n$ and $l \in \mathbb{R}^n$ we have the upper-bound:*

$$l^\top(x - y) - B_F(y\|x) \leq \frac{1}{2}\sum_{i=1}^n x(i) l(i)^2$$

Proof. Let $f_\psi(x) = -\log(x)$, then $F(x) = \sum_{i=1}^n f_\psi(x(i)) - f_\psi(1/n)$

$$l^\top(x - y) - \mathrm{B}_F(y\|x) = \sum_{i=1}^n l(i)(x(i) - y(i)) - \mathrm{B}_{f_\psi}(y(i)\|x(i))$$

Apply Corollary 2.14

$$\leq \sum_{i=1}^{n} l(i)(x(i) - y(i)) - \frac{(x(i) - y(i))^2}{2x(i)}$$
$$\leq \frac{1}{2} \sum_{i=1}^{n} l(i)^2 x(i)$$

■

Compare the result of Lemma 2.10 and Lemma 2.15. In Lemma 2.15, the inequality $l^\top(x - y) - \mathrm{B}_F(y\|x) \leq \frac{1}{2}\sum_{i=1}^{n} x(i)l(i)^2$ holds for all $l \in \mathbb{R}^n$. In Lemma 2.10, we have a slighly tighter bound $l^\top(x - y) - \mathrm{B}_F(y\|x) \leq \sum_{i=1}^{n} x(i)^2 l(i)^2$, which holds only if $\sum_{i=1}^{n} x(i)^2 l(i)^2 \leq \frac{1}{4}$. These results are summarized in Table 2.1.

Table 2.1: Upper bounds for $l^\top(x - y) - \mathrm{B}_F(y\|x)$ when $F(x) = -n\log(n) - \sum_{i=1}^{n} \log(x(i))$

Lemma	**Domain of** x, y	**Condition on** l	**Upper bound**
Lemma 2.10	$\mathbb{R}^n_+$	$\|l\|_{\nabla^2 F(x)^{-1}} \leq 1/2$	$\sum_{i=1}^{n} x(i)^2 l(i)^2$
Lemma 2.15	Δ_n	$l \in \mathbb{R}^n$	$\frac{1}{2}\sum_{i=1}^{n} x(i)l(i)^2$

2.6 Conclusion

In this chapter, we laid the technical foundations for the applications in subsequent chapters. We provided a general regret inequality for Optimistic FTRL in Theorem 2.3 from which we can obtain the regret bounds of all the algorithms in this dissertation. For tuning the learning rates in our algorithms, we use the general technique of AdaFTRL and obtain a bound on the regret using Lemma 2.6. Finally, as we use the log-barrier regularizer in some of our algorithms, we provide two techniques for bounding the stability term $\Psi_x(l)$ under different conditions. The first uses the theory of self-concordant function and is presented in Lemma 2.10. The second uses our new technique of obtaining local-norm lower bound and is presented in Lemma 2.15

Chapter 3: Scale-Free Adversarial Multi-Armed Bandits

3.1 Introduction

The Adversarial Multi Armed Bandit(MAB) problem proceeds as a sequential game of T rounds between a player and an adversary. In each round $t = 1, \ldots, T$, the player selects a distribution p_t over the n-arms and the adversary selects a loss vector l_t belonging to some set $\mathcal{L} \subseteq \mathbb{R}^n$. An action i_t is sampled from p_t and the player observes the loss $l_t(i_t)$. The (expected) regret of the player is:

$$\mathbb{E}\left[\sum_{t=1}^{T} l_t(i_t) - \min_{i \in [n]} \sum_{t=1}^{T} l_t(i)\right]$$

We assume that the adversary is oblivious, i.e., the loss vectors $l_1, \ldots, l_T$ are chosen before the game begins. So, the above expectation is with respect to the randomness in the player's strategy. The goal of the player is to sequentially select the distributions $p_1, \ldots, p_T$ such that R_T is minimized. The interaction protocol for the adversarial MAB problem is stated below:

Adversarial Multi Armed Bandit with Oblivious Adversary - Interaction Protocol:

Adversary picks the loss sequence $l_1, \ldots, l_T$
Initial information set $\mathcal{I}_1 = \{\}$
for $t = 1$ *to* T **do**
  Player picks $p_t \in \Delta_n$ and samples an action $i_t \sim p_t$
  Adversary revels $l_t(i_t)$ Player incurs cost $l_t(i_t)$
  Update information set $\mathcal{I}_{t+1} = I_t \cup \{p_t, i_t, l_t(i_t)\}$

The adversarial MAB problem has been studied extensively; we refer the reader to the texts of [34, 35, 36] for further details. Assuming that $\mathcal{L}$ is bounded, and the $\|\cdot\|_\infty$-Lipschitz constant G is known to the player in advance (i.e. $\sup_{l \in \mathcal{L}} \|l\|_\infty = G < \infty$), the minimax rate of regret is known to be $\Theta(G\sqrt{nT})$. The Exp3 algorithm [11] has a $O(G\sqrt{nT\log(n)})$ regret bound whereas

the Poly-INF algorithm [31] removes the $\sqrt{\log(n)}$ factor, achieving the optimal $O(G\sqrt{nT})$ regret bound. Exp3 and Poly-INF use G in tuning the learning rate, which helps them achieve a linear dependence on G.

In this paper, we address the case when the player has no knowledge of $\mathcal{L}$. We consider *Scale-Free* bounds for MABs, which aim to bound the regret in terms of n and norms of the loss vectors $l_1, \ldots, l_T$ for any sequence of loss vectors chosen arbitrarily by adversary. Scale-free bounds have been studied in the *full-information* setting (where the player sees the complete vector l_t in each round). For the Experts problem, which is the full-information counterpart of adversarial MAB, the AdaHedge algorithm [37] has a scale-free regret bound of $O(\sqrt{\log(n)(\sum_{t=1}^{T} \|l_t\|_\infty^2)})$. For the same problem, the Hedge algorithm [38] has a regret bound of $O(G\sqrt{T\log(n)})$ with knowledge of G. The scale-free bound is more general as it holds for any $l_1, \ldots, l_T \in \mathbb{R}^n$, whereas the bound achieved by the Hedge algorithm only holds provided that $\sup_t \|l_t\|_\infty < G$ where G needs to be known in advance.

3.1.1 Our Contributions

We present an algorithm for the scale-free MAB problem. By appropriately setting the parameters of this algorithm, we can achieve a scale-free regret upper-bound of either $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nT})$, or $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nL_1})$. Here $L_\infty = \sup_t \|l_t\|_\infty$, $L_2 = \sum_{t=1}^{T} \|l_t\|_2^2$, $L_1 = \sum_{t=1}^{T} \|l_t\|_1$ and the $\tilde{O}$ notation suppress logarithmic factors. Our algorithm is also *any-time* as it does not need to know the number of rounds T in advance. Assuming $\sup_t \|l_t\|_\infty < G$, our first regret bound achieves linear dependence on G (sans the hidden logarithmic terms). This bound is only $\tilde{O}(\sqrt{n})$ factor larger than Poly-INF's regret of $O(G\sqrt{nT})$. The second bound is the first completely data-dependent scale-free regret bound for MABs as it has no direct dependence on T. Moreover, these are the first MAB bounds that adapt to the $\|\cdot\|_2$, $\|\cdot\|_1$ norms of the losses. The only previously known scale-free result for MABs was $O(L_\infty\sqrt{nT\log(n)})$ by Hadiji and Stoltz [29], which adapts to the $\|\cdot\|_\infty$ norm and is not completely data-dependent due to the $\sqrt{T}$ dependence in their bound.

3.1.2 Related Work

Scale-Free Regret. As mentioned earlier, Scale-Free regret bounds were studied in the full information setting. The AdaHedge algorithm from Rooij *et al.* [37] gives a scale-free bound for the experts problem. The AdaFTRL algorithm from Orabona and Pál [13] extends these bounds to the general online convex optimization problem. For the MAB problem, Hadiji and Stoltz [29] show a scale-free bound of $O(L_\infty\sqrt{nT\log(n)})$, which is close to the $O(G\sqrt{nT\log(n)})$ bound of Exp3. Our scale-free bounds are more versatile as they are able to adapt to additional structure in the loss sequence, such as the case of sparse losses with large magnitude, i.e., when $L_2 << L_\infty^2 nT$ and $L_1 << L_\infty nT$. Even in the worst-case, our bounds are a factor of $\tilde{O}(\sqrt{n})$ and $\tilde{O}(\sqrt{nL_\infty})$ larger than their bound respectivley.

Data-dependent Regret. These bounds use a "measure of hardness" of the sequence of loss vectors to bound the regret of an algorithm. Algorithms that have a data-dependent regret bound perform better than the worst-case regret, when the sequence of losses is "easy" according to the measure of hardness used. For instance, First-order bounds [39, 40, 41, 42, 23], also known as small-loss or $L^\star$ bounds depend on $L^\star = \min_{i\in[n]}\sum_{t=1}^T l_t(i)$. Bounds that depend on the empirical variance of the losses were shown in [43, 44, 23, 42]. Path length bounds that depend on $\sum_{t=1}^{T-1}\|l_t - l_{t+1}\|$ or a similar quantity appear in [45, 42]. Algorithms that adapt to any stochastictiy present in the losses were given in [46, 23, 42]. Our bound is comparable to a result in [44], where they derive a regret bound depending on $\sum_{t=1}^T \|l_t\|_2^2$. However, all these results assume either $\mathcal{L} = [0,1]^n$ or $\mathcal{L} = [-1,1]^n$. We make no such assumptions in our work.

Effective Range Regret. The effective range of the loss sequence is defined as $\sup_{t,i,j}|l_t(i)-l_t(j)|$. Gerchinovitz and Lattimore [47] showed that it is impossible to adapt to the effective range in adversarial MAB. This result does not contradict the existence of scale-free bounds as the effective range could be much smaller than, for instance, the complete range $\sup_{t,s,i,j}|l_t(i) - l_s(j)|$. In fact, Hadiji and Stoltz [29] already show a regret bound that adapts to the complete range. We do

note that under some mild additional assumptions, Cesa-Bianchi and Shamir [48] show that it is possible to adapt to the effective range.

3.1.3 Notation

Let Δ_n be the probability simplex $\{p \in \mathbb{R}^n : \sum_{i=1}^n p(i) = 1, p(i) \geq 0, i \in [n]\}$. Let $\mathbf{1}^i$ be the vector with $\mathbf{1}^i(i) = 1$ and $\mathbf{1}^i(j) = 0$ for all $j \neq i$. For $\epsilon \in (0, 1]$, let $\mathbf{1}^i_\epsilon = (1 - \epsilon)\mathbf{1}^i + \epsilon/n$. The all ones and all zeros vector are denoted by $\mathbf{1}$ and $\mathbf{0}$ respectively. Let $\mathcal{I}_t$ be the history from time-step 1 to t, i.e., $\mathcal{I}_t = \{p_1, i_1, l_1(i_1), p_2, i_2, l_2(i_2), \ldots, p_t, i_t, l_t(i_t)\}$.

3.2 Algorithm

Consider for a moment, full-information strategies on Δ_n. In the full information setting, in each round t, the player picks a point $p_t \in \Delta_n$. Simultaneously, the adversary picks a loss vector $l_t \in \mathbb{R}^n$. The player incurs a loss of $l_t^\top p_t$ and (unlike the bandit setting) *sees the entire vector* l_t. A full-information strategy $\mathcal{F}$ takes as input a sequence of loss vectors $l_1, \ldots, l_t$ and outputs the next iterate $p_{t+1} \in \Delta_n$. A MAB strategy $\mathcal{B}$ can be constructed from a full-information strategy $\mathcal{F}$ along with two other components as follows:

1. A sampling scheme $\mathcal{S}$, which constructs a sampling distribution p'_t from the current iterate p_t. An arm i_t is then sampled from p'_t and the loss $l_t(i_t)$ is revealed to the player.

2. An estimation scheme $\mathcal{E}$, that constructs an estimate $\tilde{l}_t$ of the loss vector using $l_t(i_t)$ and p_t.

3. A full-information strategy $\mathcal{F}$, which computes the next iterate p_{t+1} using $\tilde{l}_1, \ldots, \tilde{l}_t$.

In fact, most existing MAB strategies in the literature can be described in the above framework with different choices of $\mathcal{S}, \mathcal{E}, \mathcal{F}$.

A delicate balance needs to be struck between $\mathcal{S}, \mathcal{E}$ and $\mathcal{F}$ in order to achieve a good regret bound for $\mathcal{B}$. Suppose the best arm in hindsight is $i_\star = \arg\min_{i \in [n]} \sum_{t=1}^T l_t(i)$ The expected regret

of MAB strategy $\mathcal{B}$ can be decomposed as follows:

$$\mathbb{E}\left[\sum_{t=1}^{T}(l_t(i_t) - l_t(i^\star))\right] = \mathbb{E}\left[\sum_{t=1}^{T} l_t^\top(p'_t - \mathbf{1}^{i^\star})\right] = \mathbb{E}\left[\sum_{t=1}^{T} l_t^\top(p'_t - p_t)\right] + \mathbb{E}\left[\sum_{t=1}^{T} l_t^\top(p_t - \mathbf{1}^{i^\star})\right]$$
$$= \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} l_t^\top(p'_t - p_t)\right]}_{(1)} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}(l_t - \tilde{l}_t^\top)(p_t - \mathbf{1}^{i^\star})\right]}_{(2)} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\tilde{l}_t^\top(p_t - \mathbf{1}^{i^\star})\right]}_{(3)}$$

Term (1) is due to the sampling scheme $\mathcal{S}$, term (2) is the effect of the estimation scheme $\mathcal{E}$ and term (3) is the expected regret of the full-information strategy $\mathcal{F}$ on the loss sequence $\tilde{l}_1, \ldots, \tilde{l}_T$ compared to playing the fixed strategy $\mathbf{1}^{i^\star}$.

Sampling Scheme. A commonly used sampling scheme mixes p_t with the uniform distribution using a parameter γ, i.e., $p'_t = (1-\gamma)p_t + \gamma/n$. Such schemes were first introduced in the seminal work of Auer *et al.* [11] and have remained a mainstay in MAB algorithm design. We use a time-varying γ, i.e., we pick $p'_t = (1-\gamma_{t-1})p_t + \gamma_{t-1}/n$.

Estimation Scheme. We use the *Importance Weighted*(IW) estimator which was also introduced by Auer *et al.* [11]. It computes $\tilde{l}_t$ as:

$$\tilde{l}_t = \frac{l_t(i_t)}{p'_t(i_t)}\mathbf{1}^{i_t}$$

Since the sampling distribution is p'_t, the IW estimator is an unbiased estimate of l_t:

$$\mathbb{E}_{i_t \sim p'_t}[\tilde{l}_t] = \sum_{i_t=1}^{n} p'_t(i_t)\frac{l_t(i_t)}{p'_t(i_t)}\mathbf{1}^{i_t} = l_t$$

Note that p_t is a measurable function of H_{t-1}. Using the tower rule and the fact that $\mathbb{E}_{i_t \sim p'_t}[\tilde{l}_t] = l_t$, we can see that term (2) is 0.

Full-information startegy. For $\mathcal{F}$, there is a large variety of full-information algorithms that one could pick from. Most if not all of them belong to one of the two principle families of algorithms:

Follow The Regularized Leader(FTRL) or Online Mirror Descent(OMD). Further, one also has to choose a suitable *regularizer* F within these algorithms for the particular application at hand. The particular algorithm we use is FTRL with an adaptive learning rate η_t that resembles the adaptive schemes in AdaHedge [37] and AdaFTRL [13].

The regret of $\mathcal{F}$ has an component called the stability term $\Psi_p : \mathbb{R}^n \to \mathbb{R}$. In the bandit case, $\mathcal{F}$ receives the IW estimates $\tilde{l}_t$. So, it is important that the stability term be bounded with IW estimates. Without going into any technical details, we note that it is desirable to have a stability term bounded by $\Psi_p(l) \leq p^\top l^2$ as its expectation with IW estimates can be bounded.

Previous techniques to bound the stability term by $p^\top l^2$ relied on the assumptions on l, such as either $l \geq \mathbf{0}$ or $l \geq -\mathbf{1}$ (See [49, Page 6]). For arbitrary $l \in \mathbb{R}^n$, we show that it is possible to bound the stability term by $p^\top l^2$ using the *log-barrier* regularizer.

The complete algorithm for the scale-free MAB problem is described below. We give two choices for the exploration parameter γ_t. A simple non-adaptive scheme that is similar to the one in [29], where $\gamma_t \propto \frac{1}{\sqrt{t}}$ and an adaptive scheme that picks γ_t in a fashion that resembles the adaptive learning rate scheme η_t.

Our main result is the following regret bound for Algorithm 1.

Theorem 3.1. *For any $l_1, \ldots, l_T \in \mathbb{R}^n$, the expected regret of Algorithm 1 is at most:*

1. $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nT})$ *if γ_t is non-adaptive (Option 1) and $T \geq 4n$*
2. $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nL_1})$ *if γ_t is adaptive (Option 2)*

Where $L_\infty = \max_t \|l_t\|_\infty$, $L_2 = \sum_{t=1}^T \|l_t\|_2^2$, $L_1 = \sum_{t=1}^T \|l_t\|_1$.

3.3 Full-information Log-Barrier AdaFTRL on Linear Functions

The iterates of FTRL with the regularizer $F_\psi(x) = \sum_{i=1}^n [f_\psi(x(i)) - f_\psi(1/n)]$ for some potential function ψ, positive non-increasing learning rates $\{\eta_t\}_{t=0}^T$ and linear cost l_t are of the form:

$$p_{t+1} = \arg\min_{q \in \Delta_n} \left[\frac{F_\psi(q)}{\eta_t} + \sum_{s=1}^{t} l_s^\top q \right] \tag{3.1}$$

Algorithm 1: Scale-Free Multi Armed Bandit

Starting Parameters: $\eta_0 = n, \gamma_0 = 1/2$

Regularizer $F(q) = \sum_{i=1}^{n} (f(q(i)) - f(1/n))$, where $f(x) = -\log(x)$

First iterate $p_1 = (1/n, \ldots, 1/n)$

for $t = 1$ *to* T **do**

Sampling Scheme: $p'_t = (1 - \gamma_{t-1})p_t + \frac{\gamma_{t-1}}{n}$

Sample Arm $i_t \sim p'_t$ and see loss $l_t(i_t)$.

Estimation Scheme: $\tilde{l}_t = \frac{l_t(i_t)}{p'_t(i_t)} \mathbf{1}^{i_t}$

Compute γ_t for next step:

(Option 1) Non-adaptive $\gamma_t = \min(1/2, \sqrt{n/t})$

(Option 2) Adaptive $\gamma_t = \frac{n}{2n + \sum_{s=1}^{t} \Gamma_s(\gamma_{s-1})}$ where $\Gamma_t(\gamma) = \frac{\gamma |l_t(i_t)|}{(1-\gamma)p_t(i_t) + \gamma/n}$

Compute $\eta_t = \frac{n}{1 + \sum_{s=1}^{t} M_s(\eta_{s-1})}$ where $M_t(\eta) = \sup_{q \in \Delta_n} \left[\tilde{l}_t^\top (p_t - q) - \frac{1}{\eta} \mathrm{B}_F(q \| p_t) \right]$

Find next iterate using FTRL: $p_{t+1} = \arg\min_{q \in \Delta_n} \left[F(q) + \eta_t \sum_{s=1}^{t} q^\top \tilde{l}_s \right]$

The point p_{t+1} always exists strictly inside Δ_n.

Corollary 3.2. *For any $l_1, \ldots, l_t \in R^n$, the iterates of Equation 3.1 satisfy the inequality:*

$$\sum_{t=1}^{T} l_t^\top (p_t - p) \leq \frac{F_\psi(p)}{\eta_T} + \sum_{t=1}^{T} l_t^\top (p_t - p_{t+1}) - \frac{1}{\eta_{t-1}} B_{F_\psi}(p_{t+1} \| p_t)$$

Proof. We apply Corollary 2.4 with regularization $F\psi(x)$, which is always non-negative and has minimum value 0. The learning rates η_t are non-increasing. Finally, as the costs are linear, the Bregman Divergence $\mathrm{B}_{g_t}(p_{t+1} \| p_t) = 0$. Thus, we get the regret inequality of the above form. ■

Define $M_t(\eta)$ as:

$$M_t(\eta_{t-1}) = \sup_{q \in \Delta_n} \left[l_t^\top (p_t - q) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(q \| p_t) \right] \geq l_t^\top (p_t - p_{t+1}) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(p_{t+1} \| p_t)$$

The AdaFTRL strategy picks an adaptive learning rate:

$$\eta_t = \frac{\alpha}{\beta + \sum_{s=1}^{t} M_s(\eta_{s-1})}$$

Where $\alpha, \beta > 0$. Since $q = p_t$ is a feasible solution for this optimization problem in the definition of $M_t(\eta_{t-1})$, we have $M_t(\eta_{t-1}) \geq 0$. Let $p_t^\star$ be the optimal value of q in the optimization. We have the upper bound

$$M_t(\eta_{t-1}) = l_t^\top (p_t - p_t^\star) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(p_t^\star \| p_t) \leq l_t^\top (p_t - p_t^\star) \leq 2\|l_t\|_\infty$$

Since $M_t(\eta_{t-1})$ are non-negative and bounded, the sequence η_t is non-increasing.

Theorem 3.3. *If the regularizer is the log-barrier $F_\psi(x) = \sum_{i=1}^{n} [\log(1/n) - \log(x(i))]$ then for any $i \in [n]$, $\epsilon \in (0, 1]$ and any sequence of losses $l_1, \ldots, l_T$, the iterates of Full-information Log-Barrier AdaFTRL on Linear Functions satisfy the regret inequality $\sum_{t=1}^{T} l_t^\top (p_t - \boldsymbol{1}_\epsilon^i)$:*

$$\leq n \log(1/\epsilon) \left(\frac{\beta}{\alpha} + \frac{2 \sup_t \|l_t\|_\infty}{\alpha} \right) + 2 \sup_t \|l_t\|_\infty + \sqrt{\sum_{t=1}^{T} p_t^\top l_t^2} \left(\frac{n \log(1/\epsilon)}{\sqrt{\alpha}} + \sqrt{\alpha} \right)$$

Proof. The log-barrier regularizer $F_\psi(x) = \sum_{i=1}^{n} [\log(1/n) - \log(x(i))]$ is obtained by using the potential $\psi(u) = -1/u$ on the domain $(-\infty, 0)$. Using Corollary 2.14, we have the lower-bound:

$$\mathrm{B}_{F_\psi}(p_t^\star \| p_t) = \sum_{i=1}^{n} \mathrm{B}_{f_\psi}(p_t^\star(i) \| p_t(i)) \geq \sum_{i=1}^{n} \frac{1}{2} \frac{(p_t(i) - p_t^\star(i))^2}{p_t(i)}$$

This gives us the upper-bound:

$$\begin{aligned} M_t(\eta_{t-1}) = l_t^\top (p_t - p_t^\star) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(p_t^\star \| p_t) &\leq \sum_{i=1}^{n} \left[l_t(i)(p_t(i) - p_t^\star(i)) - \frac{(p_t(i) - p_t^\star(i))^2}{2\eta_{t-1} p_t(i)} \right] \\ &\leq \sum_{i=1}^{n} \sup_{s \in \mathbb{R}} \left[l_t(i) s - \frac{1}{2\eta_{t-1}} \frac{s^2}{p_t(i)} \right] \leq \frac{\eta_{t-1}}{2} \sum_{n=1}^{n} p_t(i) l_t(i)^2 = \frac{\eta_{t-1}}{2} p_t^\top l_t^2 \end{aligned}$$

Thus, we have

$$\frac{M_t(\eta_{t-1})}{\eta_{t-1}} \leq \frac{1}{2} p_t^\top l_t^2$$

Applying Lemma 2.6, for any $i \in [n]$ and $\epsilon \in (0, 1]$ we have that $\sum_{t=1}^T l_t(p_t - \mathbf{1}_\epsilon^i)$:

$$\leq F_\psi(\mathbf{1}_\epsilon^i)\left(\frac{\beta}{\alpha} + \frac{2\sup_t \|l_t\|_\infty}{\alpha}\right) + 2\sup_t \|l_t\|_\infty + \sqrt{\sum_{t=1}^T p_t^\top l_t^2}\left(\frac{F_\psi(\mathbf{1}_\epsilon^i)}{\sqrt{\alpha}} + \sqrt{\alpha}\right)$$

The term $F_\psi(\mathbf{1}_\epsilon^i)$ can be bounded as:

$$\begin{aligned} F_\psi(\mathbf{1}_\epsilon^i) &= n\log(1/n) - (n-1)\log(\epsilon/n) - \log((1-\epsilon) + \epsilon/n) \\ &\leq n\log(1/n) - n\log(\epsilon/n) = n\log(1/\epsilon) \end{aligned}$$

■

Table 3.1: Technique to obtain the bound $l^\top(p-q) - \mathrm{B}_{F_\psi}(q\|p) \leq cp^\top l^2$

Result	**Condition on l**	**Regularizer F_ψ**	**Bound**
Lattimore and Szepesvári [49, Eq. 6]	$l \in [-1, \infty)^n$	Negative Entropy	$\sum_{i=1}^n p(i)l(i)^2$
Lattimore and Szepesvári [49, Eq. 6]	$l \in \mathbb{R}_n^+$	Negative-Entropy	$\frac{1}{2}\sum_{i=1}^n p(i)l(i)^2$
Lemma 2.15	$l \in \mathbb{R}^n$	Logarithmic-Barrier	$\frac{1}{2}\sum_{i=1}^n p(i)l(i)^2$

For $p \in \Delta_n$ and regularizer F_ψ, the stability term is defined as $\Psi_p(l) = \sup_{q\in\Delta_n}\left[l^\top(p-q) - \mathrm{B}_{F_\psi}(q\|p)\right]$ Observe that $\eta M_t(\eta) = \Psi_{p_t}(\eta l_t)$. For the log-barrier regularizer, we have $M_t(\eta) \leq \eta p_t^\top l_t^2$. Thus, $\Psi_p(l) \leq p^\top l^2$ for all $l \in \mathbb{R}^n$. Previously, the only known way to achieve $\Psi_p(l) \leq p^\top l^2$ was by using the negative-entropy regularizer along with the assumption $l \geq -\mathbf{1}$ (See [49, Eq. 6] or [35, Eq. 37.15]). Thus, the log-barrier regularizer gives us the same stability bound $\Psi_p(l) \leq p^\top l^2$, but it holds for all $l \in \mathbb{R}^n$. We summarize the two ways of obtaining a stability of $p^\top l^2$ in Table 3.1.

3.4 Scale-free bandit regret bounds

Theorem 3.1. *For any $l_1, \ldots, l_T \in \mathbb{R}^n$, the expected regret of Algorithm 1 is at most:*

1. $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nT})$ *if γ_t is non-adaptive (Option 1) and $T \geq 4n$*
2. $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nL_1})$ *if γ_t is adaptive (Option 2)*

Where $L_\infty = \max_t \|l_t\|_\infty$, $L_2 = \sum_{t=1}^T \|l_t\|_2^2$, $L_1 = \sum_{t=1}^T \|l_t\|_1$.

Proof. Suppose the best arm in hindsight is $i_\star = \arg\min_{i\in[n]} \sum_{t=1}^T l_t(i)$. Let $\mathbf{1}^{i_\star}$ be the vector with $\mathbf{1}^{i_\star}(i_\star) = 1$ and $\mathbf{1}^{i_\star}(i) = 0$ for all $i \neq i_\star$. Let $\mathbf{1}_\epsilon^{i_\star} = (1-\epsilon)\mathbf{1}^{i_\star} + \epsilon/n$. The exptected regret of Algorithm 1 is:

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^T l_t(i_t) - l_t(i^\star)\right] &= \mathbb{E}\left[\sum_{t=1}^T l_t^\top(p_t' - \mathbf{1}^{i_\star})\right] = \mathbb{E}\left[\sum_{t=1}^T l_t^\top(\mathbf{1}_\epsilon^{i_\star} - \mathbf{1}^{i_\star})\right] + \mathbb{E}\left[\sum_{t=1}^T l_t^\top(p_t' - \mathbf{1}_\epsilon^{i_\star})\right] \\ &= \underbrace{\mathbb{E}\left[\sum_{t=1}^T l_t^\top(\mathbf{1}_\epsilon^{i_\star} - \mathbf{1}^{i_\star})\right]}_{(1)} + \underbrace{\mathbb{E}\left[\sum_{t=1}^T l_t^\top(p_t - \mathbf{1}_\epsilon^{i_\star})\right]}_{(2)} + \underbrace{\mathbb{E}\left[\sum_{t=1}^T l_t^\top(p_t' - p_t)\right]}_{(3)}\end{aligned}$$

For term (1), we have:

$$\mathbb{E}\left[\sum_{t=1}^T l_t^\top(\mathbf{1}_\epsilon^{i_\star} - \mathbf{1}^{i_\star})\right] = \sum_{t=1}^T l_t^\top(\mathbf{1}_\epsilon^{i_\star} - \mathbf{1}^{i_\star}) \leq 2\epsilon\left\|\sum_{t=1}^T l_t\right\|_\infty = 2\epsilon S_\infty$$

For term (2), we use the fact that $\mathbb{E}[\tilde{l}_t] = l_t$:

$$\mathbb{E}\left[\sum_{t=1}^T l_t^\top(p_t - \mathbf{1}_\epsilon^{i_\star})\right] = \mathbb{E}\left[\sum_{t=1}^T \tilde{l}_t^\top(p_t - \mathbf{1}_\epsilon^{i_\star})\right]$$

Since Algorithm 1 runs log-barrier regularized AdaFTRL with the loss sequence $\tilde{l}_1, \ldots, \tilde{l}_T$, we can bound the sum inside the expectation using Theorem 3.3 as $\sum_{t=1}^T \tilde{l}_t^\top(p_t - \mathbf{1}_\epsilon^{i_\star})$:

$$\leq \log(1/\epsilon)\left(1 + 2\sup_t \|\tilde{l}_t\|_\infty\right) + 2\sup_t \|\tilde{l}_t\|_\infty + \sqrt{n\sum_{t=1}^T p_t^\top \tilde{l}_t^2}\,(\log(1/\epsilon) + 1) \qquad (*)$$

Consider the term $\sup_t \|\tilde{l}_t\|_\infty$:

$$\sup_t \|\tilde{l}_t\|_\infty = \sup_t \frac{|l_t(i_t)|}{p'_t(i_t)} = \sup_t \frac{|l_t(i_t)|}{(1-\gamma_{t-1})p_t(i_t) + \gamma_{t-1}/n} \leq n \sup_t \frac{|l_t(i_t)|}{\gamma_{t-1}}$$

Since γ_t is a positive, non-increasing sequence:

$$\sup_t \|\tilde{l}_t\|_\infty \leq n\frac{\sup_t |l_t(i_t)|}{\gamma_T} \leq \frac{nL_\infty}{\gamma_T}$$

Finally, consider the term $p_t^\top \tilde{l}_t^2$:

$$p_t^\top \tilde{l}_t^2 = p_t(i_t)\frac{l_t(i_t)^2}{p'_t(i_t)^2} = p_t(i_t)\frac{l_t(i_t)^2}{((1-\gamma_{t-1})p_t(i_t) + \frac{\gamma_{t-1}}{n})p'_t(i_t)} \leq \frac{l_t(i_t)^2}{(1-\gamma_{t-1})p'_t(i_t)}$$

Since $0 \leq \gamma_{t-1} \leq 1/2$, we have $1 \leq (1-\gamma_{t-1})^{-1} \leq 2$. Thus:

$$p_t^\top \tilde{l}_t^2 \leq 2\frac{l_t(i_t)^2}{p'_t(i_t)}$$

Substituting these bounds in the regret inequality $(*)$, we have $\sum_{t=1}^T \tilde{l}_t^\top (p_t - \mathbf{1}_\epsilon^{i_\star})$:

$$\leq \log(1/\epsilon) + \sqrt{2n\sum_{t=1}^T \frac{l_t(i_t)^2}{p'_t(i_t)}}\,(\log(1/\epsilon)+1) + \frac{2nL_\infty}{\gamma_T}\,(\log(1/\epsilon)+1)$$

Applying expectation, we have $\mathbb{E}\left[\sum_{t=1}^T \tilde{l}_t^\top (p_t - \mathbf{1}_\epsilon^{i_\star})\right]$:

$$\leq \log(1/\epsilon) + \mathbb{E}\left[\sqrt{2n\sum_{t=1}^T \frac{l_t(i_t)^2}{p'_t(i_t)}}\right](\log(1/\epsilon)+1) + 2nL_\infty\,(\log(1/\epsilon)+1)\,\mathbb{E}\left[\frac{1}{\gamma_T}\right]$$

For the expectation in the second term, we apply Jensen's inequality:

$$\mathbb{E}\left[\sqrt{2n\sum_{t=1}^T \frac{l_t(i_t)^2}{p'_t(i_t)}}\right] \leq \sqrt{2n\mathbb{E}\sum_{t=1}^T \left[\frac{l_t(i_t)^2}{p'_t(i_t)}\right]} = \sqrt{2n\sum_{t=1}^T\sum_{i=1}^n l_t(i)^2} = \sqrt{2nL_2}$$

Thus term (2) can be bounded as $\mathbb{E}\left[\sum_{t=1}^{T} l_t^\top (p_t - \mathbf{1}_\epsilon^{i_\star})\right]$:

$$\leq \log(1/\epsilon) + \sqrt{2nL_2}\,(\log(1/\epsilon)+1) + 2nL_\infty\,(\log(1/\epsilon)+1)\,\mathbb{E}\left[\frac{1}{\gamma_T}\right]$$

3.4.1 Non-Adaptive Exploration

First, we present a simple way to bound term (3):

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} l_t^\top (p_t' - p_t)\right] &= \mathbb{E}\left[\sum_{t=1}^{T} l_t^\top ((1-\gamma_{t-1})p_t + \gamma_{t-1}/n - p_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \gamma_{t-1} l_t^\top (1/n - p_t)\right] \\ &\leq \mathbb{E}\left[2\sum_{t=1}^{T} \gamma_{t-1}\|l_t\|_\infty\right] \leq 2L_\infty \mathbb{E}\left[\sum_{t=1}^{T}\gamma_{t-1}\right]\end{aligned}$$

Combining the upper-bounds for term (1), (2) and (3), we have $\mathbb{E}\left[\sum_{t=1}^{T} l_t(i_t) - l_t(i^\star)\right]$:

$$\leq 2\epsilon S_\infty + \log(1/\epsilon) + \sqrt{2nL_2}\,(\log(1/\epsilon)+1) + 2nL_\infty\,(\log(1/\epsilon)+1)\,\mathbb{E}\left[\frac{1}{\gamma_T}\right] + 2L_\infty\mathbb{E}\left[\sum_{t=1}^{T}\gamma_{t-1}\right]$$

Pick $\epsilon = (1+S_\infty)^{-1}$ and the exploration rate $\gamma_t = \min(1/2, \sqrt{n/t})$. If $T \geq 4n$, the regret of Algorithm 1 with non-adaptive exploration is bounded by:

$$\begin{aligned}&\leq 2 + \log(1+S_\infty) + \sqrt{2nL_2}(1+\log(1+S_\infty)) + 2L_\infty\sqrt{nT}(2+\log(1+S_\infty)) \\ &\leq (2+\log(1+S_\infty))\left(1 + \sqrt{2nL_2} + 2L_\infty\sqrt{nT}\right) \\ &= \tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nT})\end{aligned}$$

3.4.2 Adaptive Exploration

An alternate way to bound term (3) is:

$$\mathbb{E}\left[\sum_{t=1}^{T} l_t^\top (p_t' - p_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \tilde{l}_t^\top (p_t' - p_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \gamma_{t-1}\frac{l_t(i_t)}{p_t'(i_t)}(1/n - p_t(i_t))\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \gamma_{t-1} \frac{|l_t(i_t)|}{p'_t(i_t)}\right]$$

Combining the upper-bounds for term (1), (2) and (3), we have $\mathbb{E}\left[\sum_{t=1}^{T} l_t(i_t) - l_t(i^\star)\right]$:

$$\leq 2\epsilon S_\infty + \log(1/\epsilon) + \sqrt{2nL_2}\left(\log(1/\epsilon) + 1\right) + \mathbb{E}\left[\frac{2nL_\infty\left(\log(1/\epsilon) + 1\right)}{\gamma_T} + \sum_{t=1}^{T} \gamma_{t-1} \frac{|l_t(i_t)|}{p'_t(i_t)}\right]$$

Consider the expression inside the expectation. Let

$$\Gamma_t(\gamma) = \frac{\gamma |l_t(i_t)|}{(1-\gamma)p_t(i_t) + \gamma/n}$$

When $0 \leq \gamma \leq 1/2$, we have $0 \leq \Gamma_t(\gamma) \leq n|l_t(i_t)| \leq nL_\infty$. Moreover, we have

$$\frac{\Gamma_t(\gamma_{t-1})}{\gamma_{t-1}} = \frac{|l_t(i_t)|}{p'_t(i_t)}$$

Pick

$$\gamma_t = \frac{n}{2n + \sum_{s=1}^{t} \Gamma_s(\gamma_{s-1})}$$

We satisfy $0 \leq \gamma_t \leq 1/2$. Applying Lemma 2.6, we have:

$$\mathbb{E}\left[\frac{2nL_\infty\left(\log(1/\epsilon) + 1\right)}{\gamma_T} + \sum_{t=1}^{T} \gamma_{t-1} \frac{|l_t(i_t)|}{p'_t(i_t)}\right] = \mathbb{E}\left[\frac{2nL_\infty\left(\log(1/\epsilon) + 1\right)}{\gamma_T} + \sum_{t=1}^{T} \Gamma_t(\gamma_{t-1})\right]$$

$$\leq 2nL_\infty(2 + L_\infty)\left(\log(1/\epsilon) + 1\right) + nL_\infty + \left(2L_\infty\left(\log(1/\epsilon) + 1\right) + 1\right)\mathbb{E}\left[\sqrt{2n\sum_{t=1}^{T} \frac{|l_t(i_t)|}{p'_t(i_t)}}\right]$$

For the expectation above, we apply Jensen's inequality:

$$\mathbb{E}\left[\sqrt{2n\sum_{t=1}^{T} \frac{|l_t(i_t)|}{p'_t(i_t)}}\right] \leq \sqrt{2n\mathbb{E}\sum_{t=1}^{T}\left[\frac{|l_t(i_t)|}{p'_t(i_t)}\right]} = \sqrt{2n\sum_{t=1}^{T}\sum_{i=1}^{n}|l_t(i)|} = \sqrt{2nL_1}$$

Pick $\epsilon = (1 + S_\infty)^{-1}$. The regret of Algorithm 1 with adaptive exploration is bounded by:

$$
\begin{aligned}
&\leq 2 + \log(1 + S_\infty) + \sqrt{2nL_2}\,(\log(1 + S_\infty) + 1) \\
&\quad + 2nL_\infty(2 + L_\infty)\,(\log(1 + S_\infty) + 1) + nL_\infty + (2L_\infty\,(\log(1 + S_\infty) + 1) + 1)\,\sqrt{2nL_1} \\
&= \tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nL_1})
\end{aligned}
$$

■

3.5 Conclusion

In this chapter, we studied the adversarial MAB problem where the loss vectors could be completely arbitrary. This is a departure from the standard MAB setting in prior works, which typically assume that the losses are from the domain $[0, 1]^n$ or $[-1, 1]^n$. We provide a novel algorithm based on Log-Barrier FTRL and show that it has data-dependent regret bounds. With a non-adaptive exploration, we obtain a bound of $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nT})$ and with an adaptive exploration, we get $\tilde{O}(\sqrt{nL_2} + L_\infty\sqrt{nL_1})$. However, these bounds could be further improved. Specifically, there is a gap of $\sqrt{n}$ in the first bound and a gap of $\sqrt{nL_\infty}$ in the second bound.

Table 3.2: Comparison with Chen and Zhang [50]

Algorithm	$\tilde{O}$ **Regret**
Algorithm 1 with Non-Adaptive Exploration	$L_\infty\sqrt{nT} + \sqrt{n\sum_{t=1}^{T}\|l_t\|_2^2}$
Algorithm 1 with Adaptive Exploration	$L_\infty\sqrt{n\sum_{t=1}^{T}\|l_t\|_1} + \sqrt{n\sum_{t=1}^{T}\|l_t\|_2^2}$
Chen and Zhang [50] with Non-Adaptive Exploration	$L_\infty^-\sqrt{nT} + \sqrt{n\sum_{t=1}^{T}\|l_t\|_\infty^2}$
Chen and Zhang [50] with Adaptive Exploration	$L_\infty^-\sqrt{n\sum_{t=1}^{T}\|l_t\|_\infty} + \sqrt{n\sum_{t=1}^{T}\|l_t\|_\infty^2}$

After the publication of our work [51], the scale-free adversarial MAB problem was later studied by Chen and Zhang [50]. Their algorithm is also based on the Log-Barrier FTRL. The new technique they introduce is a novel loss clipping mechanism, using which they are able to close

the $\sqrt{n}$ gap in our bounds. We compare these results in Table 3.2. Note that in the bounds of Chen and Zhang [50], the L_∞^- term is the magnitude of the most negative entry of the losses, $L_\infty^- = \max_{t,k} |\min(l_t(k), 0)|$.

This chapter studies expected regret bounds. The study of high probability scale-free regret bounds remains an open problem.

Chapter 4: Data-Dependent Regret Bounds for Online Portfolio Selection

4.1 Introduction

The online portfolio selection problem (OPS), as formulated by Cover [1], is a repeated game of sequential investment between an investor (the player) and a market (the environment) consisting of n assets (stocks). The investor starts off with 1 unit of wealth. At the start of each investment period, indexed by $t = 1, 2, \ldots, T$, the investor distributes her wealth among the n assets according to a *portfolio* vector w_t. We limit the investor to long-only portfolios, i.e., no short selling is allowed. In this case, the portfolio vector w_t will belong to the unit simplex Δ_n, which is the set $\{w \in \mathbb{R}^n : \sum_{i=1}^n w(i) = 1, w(i) \geq 0, i \in [n]\}$. At the end of the investment period, the investor observes the *returns* (the ratios of the closing and opening prices in this period) $r_t \in \mathbb{R}^n_+$ from the market. The investor's wealth changes by a multiplicative factor of $r_t^\top w_t$. Therefore, the wealth after T periods will be $\prod_{t=1}^T (r_t^\top w_t)$. The interaction protocol for the online portfolio selection problem is stated below:

Online Portfolio Selection - Interaction Protocol:

Initial information set $\mathcal{I}_1 = \{\}$

for $t = 1$ *to* T **do**

- Investor picks the portfolio $w_t \in \Delta_n$
- Market revels $r_t \in \mathbb{R}^n$
- Investor's wealth grows by a multiplicative factor of $r_t^\top w_t$
- Update information set $\mathcal{I}_{t+1} = I_t \cup \{w_t, r_t\}$

The wealth of an investor that always selects the same portfolio w is $\prod_{t=1}^T (r_t^\top w)$. We can compare the difference in the log-wealth between the two investors:

$$\log\left(\prod_{t=1}^T (r_t^\top w)\right) - \log\left(\prod_{t=1}^T (r_t^\top w_t)\right) = \sum_{t=1}^T (-\log(r_t^\top w_t)) - (-\log(r_t^\top w))$$

Suppose we define $f_t(w) = -\log(r_t^\top w)$ as our cost function, then the above expression exactly corresponds to the notion of static regret $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$. The cost functions f_t are convex in w. Thus we can study the online portfolio selection problem as an instance of Online Convex Optimization.

4.1.1 Notation

Let Δ_n be the probability simplex $\{w \in \mathbb{R}^n : \sum_{i=1}^{n} w(i) = 1, w(i) \geq 0, i \in [n]\}$. The all ones vector are denoted by $\mathbf{1}$. The set of real numbers is $\mathbb{R}$, non-negative numbers is $\mathbb{R}_+$ and positive numbers is $\mathbb{R}_{++}$. The Hadamard product of two vectors u, v is represented as $u \circ v$.

4.2 Prior Works and Our Contributions

4.2.1 Worst-Case Regret Bounds

In his seminal work, Cover [1] proposes the OPS problem and provides a no-regret algorithm called the Universal Portfolio (UP). Later, Cover and Ordentlich [52] showed that the minimax regret for OPS is $\Theta(n \log T)$ and the Universal Portfolio obtains this rate, i.e., for any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, the regret of Cover's UP is $O(n \log T)$. The iterates of UP are:

$$w_t = \frac{\int_{\Delta_n} w \prod_{s=1}^{t-1} (r_s^\top w) dw}{\int_{\Delta_n} \prod_{s=1}^{t-1} (r_s^\top w) dw}$$

The UP updated can be interpreted as a continuous exponential weights over the probability simplex, with loss functions $f_t(w) = -\log(r_t^\top w)$.

Helmbold *et al.* [53] identify that the OPS can be posed as an OCO problem and propose using the Exponentiated Gradient (EG) [18] algorithm to solve it. EG can be interpreted as an FTRL with linear surrogate function $\tilde{f}_t(w) = f_t(w_t) + \nabla f_t(w_t)^\top (w - w_t)$, negative entropy regularization

and parameter η. It's iterates can be written as:

$$w_t = \arg\min_{w \in \Delta_n} \left(\sum_{s=1}^{t-1} -\frac{r_s}{r_s^\top w_s} \right)^\top w + \frac{1}{\eta} \sum_{i=1}^{n} w(i) \log w(i)$$

Since the EG algorithm requires the gradient of f_t to be bounded, it results in a regret bound that is not uniform over the sequence of returns. Their result states that for any sequence of returns $r_1, \ldots, r_t \in [c, C]^n$, the regret of EG with a suitable value of η is $O(C/c\sqrt{T \log n})$. Here $c, C \in \mathbb{R}^n_{++}$ are assumed to be known apriori by the investor as they are used to tune η within the EG algorithm. Helmbold *et al.* [53] also propose a technique to "universalize" EG, resulting in the $\widetilde{\text{EG}}$ algorithm, that has a regret of $O(n^{1/2}T^{3/4})$ for any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}^n_+$. This bound was later improved to $O(n^{1/3}T^{2/3})$ in [54].

While the worst-case regret of EG is worse than UP, its per iteration run time is $O(n)$. On the other hand, the fastest known implementation of UP requires a per iteration run time of $O(n^4T^{14})$ as it involves approximating a high dimensional integral via log-concave sampling [55].

Hazan *et al.* [3] propose the Online Newton Step(ONS) method. It can be interpreted as an FTRL with a quadratic surrogate $\tilde{f}_t(w) = f_t(w_t) + \nabla f_t(w_t)^\top (w - w_t) + \frac{\beta}{2}(\nabla f_t(w_t)^\top (w - w_t))^2$ and an ℓ_2 regularizer . The iterates of ONS can be written as:

$$w_t \in \arg\min_{w \in \Delta_n} \left(\sum_{s=1}^{t-1} -\frac{r_s}{r_s^\top w_s} \right)^\top w + \frac{1}{2} w^\top \left(a \sum_{s=1}^{t-1} \frac{r_s r_s^\top}{(r_s^\top w_s)^2} + bI_n \right) w \tag{4.1}$$

Here $a, b \geq 0$ are parameters. When $b = 0$, ONS is also called Follow The Approximate Leader (FTAL) [3]. For suitable values of the parameters, ONS has $O\left(\frac{C}{c} n \log T\right)$ regret for all sequences of returns in $[c, C]^n$ and a run time of $O(n^3)$ per iteration. $c, C \in \mathbb{R}^n_{++}$ are assumed to be known by the investor as they are used to tune a, b within the ONS. Using the universalization technique [53], it is possible to modify ONS to obtain a regret of $O(n\sqrt{T \log T})$ for all return sequences in $\mathbb{R}^n_+$. The universalized version of ONS is $\widetilde{\text{ONS}}$.

Since then, there have been several works that explore the Pareto frontier of the worst case regret-run time. These include Soft-Bayes [56], AdaBARRONS [57], BISONS [58], PAE+DONS

[59], LB-OMD[54] and VB-FTRL[60]. Of these, VB-FTRL has $O(n \log T)$ worst-case regret with runtime $O(n^2 T)$, thus matching the regret of Cover's UP with much better run time. We summarize these algorithms in Table 4.1

Table 4.1: Worst-case Regret Bounds for Online Portfolio Selection

Algorithm	**Worst-case Regret**	**Run-time**	**Returns Domain**
Cover [1, 55]	$n \log(T)$	$n^4 T^{14}$	$\mathbb{R}^n_+$
EG [53]	$(C/c)\sqrt{T \log n}$	n	$[c, C]^n$, $0 < c < C$
$\widetilde{\text{EG}}$ [53, 54]	$n^{1/3} T^{2/3}$	n	$\mathbb{R}^n_+$
ONS[3]	$(C/c) n \log T$	n^3	$[c, C]^n$, $0 < c < C$
$\widetilde{\text{ONS}}$	$n\sqrt{T \log T}$	n^3	$\mathbb{R}^n_+$
Soft-Bayes [56]	$\sqrt{nT \log n}$	n	$\mathbb{R}^n_+$
AdaBARRONS [57]	$n^2 \log^4 T$	$n^{2.5} T$	$\mathbb{R}^n_+$
BISONS [58]	$n^2 \log^2 T$	n^3	$\mathbb{R}^n_+$
PAE+DONS [59]	$n^2 \log^5 T$	n^3	$\mathbb{R}^n_+$
LB-OMD[54]	$\sqrt{nT \log T}$	n	$\mathbb{R}^n_+$
VB-FTRL[60]	$n \log T$	$n^2 T$	$\mathbb{R}^n_+$

4.2.2 Data-dependent Regret Bounds

In this chapter, we aim to study algorithms with data-dependent regret bounds for OPS. Possibly the most simple algorithm for OPS is Follow-The-Leader (FTL). The iterates of FTL are obtained as:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} -\log(r_s^\top w)$$

Agarwal and Hazan [61] introduce the Smooth Prediction (SP) algorithm that adds the log-barrier regularizer to FTL. SP's iterates are obtained as:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} -\log(r_s^\top w) + \sum_{i=1}^{n} -\log(w(i))$$

A similar algorithm called Exp-Concave FTL [62], uses the ℓ_2-regularizer instead of the log-barrier:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} -\log(r_s^\top w) + \frac{1}{2}\|w\|_2^2$$

Notice that there are no parameters to tune in these algorithms. All three of these can be shown to possess the following data-dependent regret bound. For any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}^n_+$, the regret of FTL, SP, and Exp-Concave FTL is $O(R^2 n \log T)$, where $R = \max_{t,i,j} \frac{r_t(i)}{r_t(j)}$ is the data-dependent quantity. While in the worst-case, R could be unbounded, it nevertheless yields reasonable regret bounds for benign sequences of returns $r_1, \ldots, r_T$. The per-iteration run time of these algorithms is $O(n^{3.5}T)$.

Using an adaptive variant of EG called the AdaHedge algorithm [63, 37], one can obtain a regret of $O(R\sqrt{T \log n})$ for any sequence of returns. Using so-called *universal online convex optimization* (UOCO) algorithms such as Metagrad [64] and Maler [65], we can obtain a regret bound of $O(Rn \log T)$. However, UOCO algorithms function by running $O(\log T)$ instances of ONS instantiated with different parameters and running a meta experts algorithm to control them. Using this two-step approach, UOCO algorithms adapt to the optimal setting of parameters without knowing the range $[c, C]$ before-hand. Thus, the run-time complexity is $O(n^3 \log T)$ per iteration.

We show that a simple variant of the ONS algorithm that employs *adaptive-curvature surrogate functions* has $O(Rn \log T)$ regret and $O(n^3)$ running time. We term this algorithm AdaCurv ONS. Thus, we can avoid the use of complicated UOCO algorithms like Metagrad [64] and Maler [65]. However, all the above algorithms have unbounded regret in the worst case, due to the dependence on R. For some sequences of returns, R could be arbitrarily large. To avoid this issue, we propose adding the log-barrier regularizer along with an adaptivley tuned learning rate to AdaCurv ONS,

obtaining a regret bound of $O(\min(Rn\log T, \sqrt{nT\log T}))$. We term this algorithm LB-AdaCurv ONS. These prior works, along with our contributions are summarized in Table 4.2.

Table 4.2: Data-dependent Regret Bounds for Online Portfolio Selection

Algorithm	**Data-dependent Regret**	**Worst-case Regret**	**Run-time**
FTL	$R^2 n\log(T)$	∞	$n^{2.5}T$
SP [61]	$R^2 n\log(T)$	∞	$n^{2.5}T$
Exp-Concave FTL [62]	$R^2 n\log(T)$	∞	$n^{2.5}T$
AdaHedge [63, 37]	$R\sqrt{T\log n}$	∞	n
Metagrad [64]/ Maler [65]	$Rn\log(T)$	∞	$n^3\log T$
AdaCurv FTAL/ONS	$Rn\log(T)$	∞	n^3
LB-AdaCurv FTAL/ONS	$Rn\log(T)$	$\sqrt{nT\log T}$	n^3

4.3 AdaCurv ONS

The AdaCurv ONS algorithm is based on the ONS algorithm of Hazan *et al.* [3]. So, we first discuss a few details of ONS here for context.

4.3.1 Online Newton Step

In the OCO setting, assume the loss functions satisfy $f_1, \ldots, f_T$ the following condition for all $x, y \in \mathcal{D}$ for a known value of β:

$$f_t(x) \geq f_t(y) + \nabla f_t(y)^\top (x-y) + \frac{\beta}{2}(\nabla f_t(y)^\top (x-y))^2 \tag{4.2}$$

This property is also called *β-Directional Strong Convexity*. The ONS algorithm uses a surrogate function $\tilde{f}_t(w)$ defined as:

$$\tilde{f}_t(w) = f_t(w_t) + \nabla f_t(w_t)^\top (w - w_t) + \frac{\beta}{2}(\nabla f_t(w_t)^\top (w - w_t))^2$$

Notice that $\tilde{f}_t(w) \leq f_t(w)$ for all $w \in \mathcal{D}$ and $\tilde{f}_t(w_t) = f_t(w_t)$. Moreover, $\tilde{f}_t(w)$ is a quadratic function in w. For a parameter ϵ, the ONS update is:

$$w_t \in \arg\min_{w \in \mathcal{D}} \sum_{s=1}^{t-1} \tilde{f}_s(w) + \frac{\epsilon}{2}\|w\|_2^2$$

When applying ONS for the OPS problem, $f_t(w) = -\log(r_t^\top w)$, $\mathcal{D} = \Delta_n$ and $\nabla f_t(w_t) = \frac{-r_t}{r_t^\top w_t}$. Using $\frac{\beta}{1+\beta} = a$ and $\frac{\epsilon}{1+\beta} = b$, we get the ONS update presented in Equation (4.1).

$$\begin{aligned}
w_t &\in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} \left[\frac{\beta}{2} \frac{(r_s^\top w)^2}{(r_s^\top w_s)^2} - (1+\beta) \frac{r_s^\top w}{r_s^\top w_s} \right] + \frac{\epsilon}{2}\|w\|_2^2 \\
&\in \arg\min_{w \in \Delta_n} \left(\sum_{s=1}^{t-1} -\frac{r_s}{r_s^\top w_s} \right)^\top w + \frac{1}{2} w^\top \left(\frac{\beta}{1+\beta} \sum_{s=1}^{t-1} \frac{r_s r_s^\top}{(r_s^\top w_s)^2} + \frac{\epsilon}{1+\beta} I_n \right) w \\
&\in \arg\min_{w \in \Delta_n} \left(\sum_{s=1}^{t-1} -\frac{r_s}{r_s^\top w_s} \right)^\top w + \frac{1}{2} w^\top \left(a \sum_{s=1}^{t-1} \frac{r_s r_s^\top}{(r_s^\top w_s)^2} + b I_n \right) w
\end{aligned}$$

We follow the analysis of ONS and OPS from Orabona [8, Section 7.10, Section 12.5]. Using the *exp-concavity* property, we can show that when $r_1, \ldots, r_T \in [c, C]^n$ for some $0 < c < C$, then $-\log(r_t^\top w)$ satisfies Equation (4.2) with $\beta = \frac{c}{4C}$, i.e., the following inequality holds for $x, y \in \Delta_n$:

$$-\log(r_t^\top x) \geq -\log(r_t^\top y) - \frac{r_t^\top (x - y)}{r_t^\top y} + \frac{c}{8C} \left(\frac{r_t^\top (x - y)}{r_t^\top y} \right)^2 \tag{4.3}$$

This leads to a regret bound of:

$$\frac{\epsilon}{2} + \frac{2nC}{c} \log\left(1 + \frac{CT}{4c\epsilon}\right)$$

Finally, choosing $\epsilon = \frac{C}{c}$, we get the $O((C/c) n \log T)$ bound for ONS.

We seek to obtain a data-dependent regret bound by first using an adaptive curvature surrogate function instead of the constant curvature surrogate function above. Next, instead of an ϵ that needs to be tuned, we set ϵ to be a universal constant (like 1 or 0).

4.3.2 New Adaptive Curvature Surrogate Function for $-\log(r_t^\top w)$

First, we show that $f_t(w) = -\log(r_t^\top w)$ satisfies a directional strong convexity condition, but with a β that depends on the point y and r_t.

Lemma 4.1. *For all $x, y \in \Delta_n, r_t \in \mathbb{R}^n_+$ such that $r_t^\top x, r_t^\top y > 0$, we have the inequality:*

$$-\log(r_t^\top x) \geq -\log(r_t^\top y) - \frac{r_t^\top (x-y)}{r_t^\top y} + \frac{r_t^\top y}{2\max_i r_t(i)} \left(\frac{r_t^\top (x-y)}{r_t^\top y}\right)^2$$

Proof. When $r_t^\top x, r_t^\top y > 0$, we have $0 < \frac{r_t^\top x}{\max_i r_t(i)}, \frac{r_t^\top y}{\max_i r_t(i)} \leq 1$. We apply Corollary 2.14 (with $x = \frac{r_t^\top y}{\max_i r_t(i)}$ and $y = \frac{r_t^\top x}{\max_i r_t(i)}$) to obtain:

$$\begin{aligned}
&\frac{r_t^\top x}{r_t^\top y} - 1 - \log\left(\frac{r_t^\top x}{r_t^\top y}\right) \geq \frac{1}{2}\frac{(r_t^\top x - r_t^\top y)^2}{(\max_i r_t(i))(r_t^\top y)} \\
&\implies -\log(r_t^\top x) \geq -\log(r_t^\top y) - \frac{r_t^\top (x-y)}{r_t^\top y} + \frac{r_t^\top y}{2\max_i r_t(i)} \left(\frac{r_t^\top (x-y)}{r_t^\top y}\right)^2
\end{aligned}$$

■

Comparing the result of Lemma 4.1 with Equation (4.3), we see that $f_t(w)$ satisfies an adaptive directional strong convexity condition with $\beta_t(y) = \frac{r_t^\top y}{\max_i r_t(i)}$.

Using Lemma 4.1, we have the following adaptive curvature surrogate function:

$$\tilde{f}_t(w) = f_t(w_t) + \nabla f_t(w_t)^\top (w - w_t) + \frac{r_t^\top w_t}{2(\max_i r_t(i))} (\nabla f_t(w_t)^\top (w - w_t))^2 \tag{4.4}$$

Plugging the surrogate from Equation 4.4 into ONS, we get the Adaptive Curvature ONS

(AdaCurv ONS) update:

$$w_t \in \arg\min_{w\in\Delta_n} \left(\sum_{s=1}^{t-1} -\frac{r_s}{r_s^\top w_s} + \sum_{s=1}^{t-1} -\frac{r_s}{\max_i r_s(i)}\right)^\top w + \frac{1}{2} w^\top \left(\sum_{s=1}^{t-1} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I_n\right) w \quad (4.5)$$

The Adaptive Curvature FTAL algorithm modifies the FTAL algorithm in Hazan *et al.* [3] by using adaptive curvature surrogates. FTAL is obtained by using $\epsilon = 0$ in ONS. The AdaCurv FTAL algorithm update is:

$$w_t \in \arg\min_{w\in\Delta_n} \left(\sum_{s=1}^{t-1} -\frac{r_s}{r_s^\top w_s} + \sum_{s=1}^{t-1} -\frac{r_s}{\max_i r_s(i)}\right)^\top w + \frac{1}{2} w^\top \left(\sum_{s=1}^{t-1} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))}\right) w \quad (4.6)$$

Notice that there are no parameters to tune in the above update. Moreover, the iterates w_t are invariant to scaling of returns.

4.3.3 AdaCurv ONS Regret Bound

Theorem 4.2. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}^n_+$, define $f_t(w) = -\log(r_t^\top w)$. The updates of AdaCurv ONS (Equation* (4.5)*) satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \le \frac{\epsilon}{2} + \inf_{\lambda\ge 0}\left(\lambda T + \frac{nR}{2}\log\left(1 + \frac{\sum_{t=1}^{T}\|\hat{r}_t\|_2^2}{n(\epsilon+\lambda)}\right)\right)$$

Where $\hat{r}_t = \frac{r_t}{\sqrt{(r_t^\top w_t)(\max_i r_t(i))}} = \frac{r_t}{r_t^\top w_t}\sqrt{\frac{r_t^\top w_t}{\max_i r_t(i)}}$. If we set $\epsilon = 1$, we get the data-dependent regret bound for AdaCurv ONS (Equation (4.5) *with $\epsilon = 1$):*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \le \frac{1}{2} + \frac{nR}{2}\log(1 + TR)$$

If we set $\epsilon = 0$, we get the data-dependent regret bound for AdaCurv FTAL (Equation (4.6)*):*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \le R + \frac{nR}{2}\log\left(1 + T^2\right)$$

Proof. Recall the adaptive curvature surrogate function in Equation (4.4):

$$\tilde{f}_t(w) = f_t(w_t) + \nabla f_t(w_t)^\top (w - w_t) + \frac{r_t^\top w_t}{2(\max_i r_t(i))} (\nabla f_t(w_t)^\top (w - w_t))^2$$

Due to Lemma 4.1, we know that that $\tilde{f}_t(w) \leq f_t(w)$ for all $w \in \mathcal{D}$ and $\tilde{f}_t(w_t) = f_t(w_t)$. Thus,

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w)$$

Applying Corollary 2.4 and with the constant regularizer $\frac{\epsilon}{2}\left(\|w\|_2^2 - \frac{1}{n}\right)$, we get:

$$\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w) \leq \frac{\epsilon}{2}\|w\|_2^2 + \sum_{t=1}^{T} \nabla \tilde{f}_t(w_t)^\top (w_t - w_{t+1}) - \mathrm{B}_{\tilde{g}_t}(w_{t+1} \| w_t) - \frac{\epsilon}{2}\|w_{t+1} - w_t\|_2^2$$

Here $\tilde{g}_t = \sum_{s=1}^{t} \tilde{f}_t$. Note that $\nabla \tilde{f}_t(w_t) = \nabla f_t(w_t) = -\frac{r_t}{r_t^\top w_t}$. Since $\tilde{g}_t(w)$ is quadratic in w, we have $\mathrm{B}_{\tilde{g}_t}(w_{t+1} \| w_t)$

$$= \frac{1}{2}(w_{t+1} - w_t)^\top \nabla^2 \tilde{g}_t(w_t)(w_{t+1} - w_t) = \frac{1}{2}(w_{t+1} - w_t)^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} \right) (w_{t+1} - w_t)$$

Thus $\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq$

$$\frac{\epsilon}{2}\|w\|_2^2 + \sum_{t=1}^{T} -\frac{r_t}{r_t^\top w_t}^\top (w_t - w_{t+1}) - \frac{1}{2}(w_{t+1} - w_t)^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I \right) (w_{t+1} - w_t)$$

Add and subtract $\frac{\lambda}{2}\|w_{t+1} - w_t\|_2^2$.

$$= \frac{\epsilon}{2}\|w\|_2^2 + \sum_{t=1}^{T} -\frac{r_t}{r_t^\top w_t}^\top (w_t - w_{t+1}) - \frac{1}{2}(w_{t+1} - w_t)^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I + \lambda I \right) (w_{t+1} - w_t)$$
$$+ \sum_{t=1}^{T} \frac{\lambda}{2}\|w_{t+1} - w_t\|_2^2$$

Upper bound the quadratic function in $w_{t+1} - w_t$.

$$\leq \frac{\epsilon}{2}\|w\|_2^2 + \frac{1}{2}\sum_{t=1}^{T} \frac{r_t}{r_t^\top w_t}^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I + \lambda I\right)^{-1} \frac{r_t}{r_t^\top w_t} + \sum_{t=1}^{T} \frac{\lambda}{2}\|w_{t+1} - w_t\|_2^2$$

Since $w, w_t \in \Delta_n$, we can bound $\|w_{t+1} - w_t\|_2^2 \leq 2$ and $\|w\|_2^2 \leq 1$. Using Lemma B.2, we have:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{\epsilon}{2} + \frac{nR}{2}\log\left(1 + \frac{\sum_{t=1}^{T}\|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)}\right) + \lambda T$$

Since $\lambda \geq 0$ can be chosen arbitrarily, we have the bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{\epsilon}{2} + \inf_{\lambda \geq 0}\left(\lambda T + \frac{nR}{2}\log\left(1 + \frac{\sum_{t=1}^{T}\|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)}\right)\right)$$

Choosing $\epsilon = 1$, $\lambda = 0$ and noting that $\|\hat{r}_t\|_2^2 \leq nR$ gives the regret bound for AdaCurv ONS.

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{1}{2} + \frac{nR}{2}\log(1 + TR)$$

Setting $\epsilon = 0$ and $\lambda = R/T$ gives us the regret bound for AdaCurv FTAL.

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq R + \frac{nR}{2}\log\left(1 + T^2\right)$$

■

Consider a sequence of returns over n stocks where in any given period t and stock i, the $r_t(i) \in [c, C]$. In this case $R = C/c$. The regret of AdaCurv ONS for any sequence of such returns is bounded by $O((C/c)n\log T)$. However, AdaCurv ONS does not need to know the values of C and c beforehand to achieve this regret. Whereas the ONS algorithm of Hazan *et al.* [3] would need to known C and c to achieve the same regret bound.

Now consider a sequence of returns where $\min_i r_t(i) = 1/t$ and $\max_i r_t(i) = 1$. For such a sequence, $R = T$. Thus, the regret of AdaCurv ONS may be unbounded for such sequences.

In the next section, we present the LB-AdaCurv ONS algorithm that has a worst-case regret of $O(\sqrt{nT \log T})$ for any sequence of returns, while simultaneously achieving the data-dependent regret bound of $O(nR \log T)$.

4.4 LB-AdaCurv ONS

While AdaCurv ONS has $O(nR \log T)$ data-dependent regret, it's worst case regret is not uniformly bounded for all sequences of returns. Indeed, R could be $O(T)$ for some sequences of returns. To avoid this issue, we add extra regularization via the log-barrier regularizer. By tuning the strength of the regularization, we are able to obtain a worst-case regret bound while maintaining the data-dependent regret bound. We consider updates of the form:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} \tilde{f}_s(w) + \frac{\epsilon}{2}\|w\|_2^2 + \frac{1}{\eta_{t-1}} F_\psi(w) \tag{4.7}$$

Here $F_\psi(w) = \sum_{i=1}^n [\log(1/n) - \log(w(i))]$, is the log-barrier regularizer, $\tilde{f}_t(w)$ are adaptive curvature surrogate functions of the form in Equation (4.4). We pick the parameters η_t using the AdaFTRL technique as:

$$\eta_t = \frac{n \log T}{2n \log T + \sum_{s=1}^t M_s(\eta_{s-1})}$$

Let $\tilde{g}_t = \sum_{s=1}^t \tilde{f}_t$. Then $M_t(\eta_{t-1})$ is

$$M_t(\eta_{t-1}) = \sup_{w \in \Delta_n} \nabla f_t(w_t)^\top (w_t - w) - \mathrm{B}_{\tilde{g}_t}(w\|w_t) - \frac{\epsilon}{2}\|w - w_t\|_2^2 - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(w\|w_t)$$

The LB-AdaCurv ONS algorithm is described in Algorithm 2. The BARRONS algorithm of Luo *et al.* [57] also utilizes a log-barrier regularized ONS procedure. However, their algorithm uses the a constant-curvature surrogate function and an increasing sequence of parameters η_t. On the other hand, we use an adaptive surrogate function and a decreasing sequence of parameters η_t chosen via the AdaFTRL technique.

Algorithm 2: Log-Barrier Regularized Adaptive Curvature Online Newton Step

Input Parameter: ϵ

Starting Parameters: $\eta_0 = 1/2$

Regularizer $F_\psi(q) = \sum_{i=1}^{n}(f_\psi(w(i)) - f(1/n))$, where $f_\psi(x) = -\log(x)$

for $t = 1$ *to* T **do**

Pick portfolio:

$$w_t \in \arg\min_{w\in\Delta_n} \sum_{s=1}^{t-1} \tilde{f}_s(w) + \frac{\epsilon}{2}\|w\|_2^2 + \frac{1}{\eta_{t-1}} F_\psi(w)$$

Observe returns vector r_t. Let $f_t(w) = -\log(r_t^\top w)$

Construct adaptive curvature surrogate function

$$\tilde{f}_t(w) = f_t(w_t) + \nabla f_t(w_t)^\top (w - w_t) + \frac{r_t^\top w_t}{2(\max_i r_t(i))}(\nabla f_t(w_t)^\top (w - w_t))^2$$

Let $\tilde{g}_t = \sum_{s=1}^{t} \tilde{f}_t$. Compute $M_t(\eta_{t-1})$ by solving the optimization :

$$M_t(\eta_{t-1}) = \sup_{q\in\Delta_n} \left[\nabla f_t(w_t)^\top (w_t - w) - \mathrm{B}_{\tilde{g}_t}(w\|w_t) - \frac{\epsilon}{2}\|w - w_t\|_2^2 - \frac{1}{\eta_{t-1}}\mathrm{B}_{F_\psi}(w\|w_t)\right]$$

Compute

$$\eta_t = \frac{n\log T}{2n\log T + \sum_{s=1}^{t} M_s(\eta_{s-1})}$$

Like before, we analyze the regret of Algorithm 2 for $\epsilon = 1$ and $\epsilon = 0$. The $\epsilon = 0$ case is termed as LB-AdaCurv FTAL.

Theorem 4.3. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}^n_+$, define $f_t(w) = -\log(r_t^\top w)$. The iterates of LB-AdaCurv ONS (Algorithm 2) satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + \frac{\epsilon}{2} + 2n\log T + 2\min\left[\inf_{\lambda\geq 0}\left(\lambda T + \frac{nR}{2}\log\left(1 + \frac{\sum_{t=1}^{T}\|\hat{r}_t\|_2^2}{n(\epsilon+\lambda)}\right)\right), 1 + \sqrt{2n\left(\sum_{t=1}^{T}\inf_{c}\|(\nabla f_t(w_t) + c\mathbf{1})\circ w_t\|_2^2\right)\log(T)}\right]$$

Where $\hat{r}_t = \frac{r_t}{\sqrt{(r_t^\top w_t)(\max_i r_t(i))}} = \frac{r_t}{r_t^\top w_t}\sqrt{\frac{r_t^\top w_t}{\max_i r_t(i)}}$. If we set $\epsilon = 1$, we get the bound for LB-AdaCurv ONS:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{5}{2} + 2n\log T + \min\left(nR\log\left(1 + RT\right), 2 + 2\sqrt{2nT\log(T)}\right)$$

If we set $\epsilon = 0$, we get the bound for LB-AdaCurv FTAL:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + 2n\log(T) + \min\left(2R + nR\log\left(1 + T^2\right), 2 + 2\sqrt{2nT\log(T)}\right)$$

Proof. First, we decompose the regret into two terms:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) = \sum_{t=1}^{T} f_t(w_t) - f_t(w^\gamma) + \sum_{t=1}^{T} f_t(w^\gamma) - f_t(w)$$

Here $w^\gamma = (1-\gamma)w + \gamma/n$. For the second term, we have:

$$\sum_{t=1}^{T} f_t(w^\gamma) - f_t(w) = \sum_{t=1}^{T}\log\left(\frac{r_t^\top w}{(1-\gamma)r_t^\top w + \gamma r_t^\top \frac{1}{n}}\right) \leq T\log\left(\frac{1}{1-\gamma}\right) \leq 2\gamma T$$

Here, we used the fact that when $\gamma \leq 1/2$, we have $\log\left(\frac{1}{1-\gamma}\right) \leq 2\gamma$. For the first term, we use the surrogate function property:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w^\gamma) \leq \sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^\gamma)$$

So, we have:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^\gamma) + 2\gamma T$$

The iterates of LB-AdaCurv ONS are given by:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} \tilde{f}_s(w) + \frac{\epsilon}{2}\|w\|_2^2 + \frac{1}{\eta_{t-1}} \sum_{i=1}^{n} -\log(w(i))$$

The updates can be viewed as an FTRL with time varying learning rate η_{t-1} for regularizer $F_\psi(w) = \sum_{i=1}^{n}[\log(1/n) - \log(w(i))]$ and constant regularizer $\frac{\epsilon}{2}\left(\|w\|_2^2 - \frac{1}{n}\right)$. Here $F_\psi(w)$ is the log-barrier regularizer. Using Corollary 2.5 we have the regret bound $\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^\gamma)$

$$\leq \frac{\epsilon}{2}\|w^\gamma\|_2^2 + \frac{F_\psi(w^\gamma)}{\eta_T} + \sum_{t=1}^{T} \nabla \tilde{f}_t(w_t)^\top (w_t - w_{t+1}) - \mathrm{B}_{\tilde{g}_t}(w_{t+1}\|w_t) - \frac{\epsilon}{2}\|w_{t+1} - w_t\|_2^2 - \frac{1}{\eta_{t-1}}\mathrm{B}_{F_\psi}(w_{t+1}\|w_t)$$

Here $\tilde{g}_t = \sum_{s=1}^{t} \tilde{f}_s$. We compute $\nabla \tilde{f}_t(w_t) = \nabla f_t(w_t) = -\frac{r_t}{r_t^\top w_t}$. Moreover, $\tilde{g}_t(w)$ is quadratic in w, we have $\mathrm{B}_{\tilde{g}_t}(w_{t+1}\|w_t)$

$$= \frac{1}{2}(w_{t+1} - w_t)^\top \nabla^2 \tilde{g}_t(w_t)(w_{t+1} - w_t) = \frac{1}{2}(w_{t+1} - w_t)^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))}\right)(w_{t+1} - w_t)$$

We bound $F_\psi(w^\gamma)$ below:

$$\begin{aligned} F_\psi(w^\gamma) &= n\log(1/n) - (n-1)\log(\gamma/n) - \log((1-\gamma) + \gamma/n) \\ &\leq n\log(1/n) - n\log(\gamma/n) = n\log(1/\gamma) \end{aligned}$$

So for the first term, we have:

$$\frac{\epsilon}{2}\|w^{\gamma}\|_2^2 + \frac{F_{\psi}(w^{\gamma})}{\eta_T} \leq \frac{\epsilon}{2} + \frac{n}{\eta_T}\log\left(\frac{1}{\gamma}\right)$$

Define $M_t(\eta_{t-1})$ as

$$M_t(\eta_{t-1}) = \sup_{w\in\Delta_n} \nabla f_t(w_t)^{\top}(w_t - w) - \mathrm{B}_{\tilde{g}_t}(w\|w_t) - \frac{\epsilon}{2}\|w - w_t\|_2^2 - \frac{1}{\eta_{t-1}}\mathrm{B}_{F_{\psi}}(w\|w_t)$$

Let $w_t^{\star}$ be the optimal value of w in the optimization. We pick η_t as:

$$\eta_t = \frac{\alpha}{\beta + \sum_{s=1}^{t} M_s(\eta_{s-1})}$$

We bound the regret in in two different ways. Substituting η_t in the regret inequality, we have:

$$\begin{aligned}\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^{\gamma}) &\leq \frac{\epsilon}{2} + \frac{n}{\eta_T}\log\left(\frac{1}{\gamma}\right) + \sum_{t=1}^{T} M_t(\eta_{t-1}) \\ &= \frac{\epsilon}{2} + \frac{n\log(1/\gamma)\beta}{\alpha} + \left(\frac{n\log(1/\gamma)}{\alpha} + 1\right)\left(\sum_{t=1}^{T} M_t(\eta_{t-1})\right)\end{aligned}$$

Observe that $M_t(\eta_{t-1})$ can be written as:

$$M_t(\eta_{t-1}) = \nabla f_t(w_t)^{\top}(w_t - w_t^{\star}) - \mathrm{B}_{\tilde{g}_t}(w_t^{\star}\|w_t) - \frac{\epsilon}{2}\|w_t^{\star} - w_t\|_2^2 - \frac{1}{\eta_{t-1}}\mathrm{B}_{F_{\psi}}(w_t^{\star}\|w_t)$$

Ignoring the last Bregman term.

$$\leq \nabla f_t(w_t)^{\top}(w_t - w_t^{\star}) - \mathrm{B}_{\tilde{g}_t}(w_t^{\star}\|w_t) - \frac{\epsilon}{2}\|w_t^{\star} - w_t\|_2^2$$

Adding and subtracting $\frac{\lambda}{2}\|w_t^{\star} - w_t\|_2^2$.

$$= \nabla f_t(w_t)^{\top}(w_t - w_t^{\star}) - \mathrm{B}_{\tilde{g}_t}(w_t^{\star}\|w_t) - \frac{\epsilon + \lambda}{2}\|w_t^{\star} - w_t\|_2^2 + \frac{\lambda}{2}\|w_t^{\star} - w_t\|_2^2$$

Taking the sum, we have $\sum_{t=1}^{T} M_t(\eta_{t-1})$

$$
\begin{aligned}
&\leq \sum_{t=1}^{T} \left(\nabla f_t(w_t)^\top (w_t - w_t^\star) - \mathrm{B}_{\tilde{g}_t}(w_t^\star \| w_t) - \frac{\epsilon + \lambda}{2} \|w_t^\star - w_t\|_2^2 \right) + \sum_{t=1}^{T} \frac{\lambda}{2} \|w_t^\star - w_t\|_2^2 \\
&\leq \sum_{t=1}^{T} -\frac{r_t}{r_t^\top w_t}^\top (w_t - w_t^\star) - \frac{1}{2}(w_t^\star - w_t)^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I + \lambda I \right) (w_t^\star - w_t) \\
&\quad + \sum_{t=1}^{T} \frac{\lambda}{2} \|w_t^\star - w_t\|_2^2 \\
&\leq \frac{\lambda}{2} \sum_{t=1}^{T} \|w_t^\star - w_t\|_2^2 + \frac{1}{2} \sum_{t=1}^{T} \frac{r_t}{(r_t^\top w_t)}^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I + \lambda I \right)^{-1} \frac{r_t}{(r_t^\top w_t)}
\end{aligned}
$$

Using Lemma B.2 and using the fact that $\lambda \geq 0$ can be chosen arbitrarily, we have:

$$
\sum_{t=1}^{T} M_t(\eta_{t-1}) \leq \inf_{\lambda \geq 0} \left(\lambda T + \frac{nR}{2} \log \left(1 + \frac{\sum_{t=1}^{T} \|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)} \right) \right)
$$

Thus, we have the following bound:

$$
\frac{F(w^\gamma)}{\eta_T} + \sum_{t=1}^{T} M_t(\eta_{t-1}) \leq \frac{\epsilon}{2} + \frac{n \log(1/\gamma)\beta}{\alpha} + \left(\frac{n \log(1/\gamma)}{\alpha} + 1 \right) \left(\inf_{\lambda \geq 0} \left(\lambda T + \frac{nR}{2} \log \left(1 + \frac{\sum_{t=1}^{T} \|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)} \right) \right) \right)
$$

Using $\alpha = n \log T$, $\beta = 2n \log T$ and $\gamma = 1/T$, the above bound yields:

$$
\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^\gamma) \leq \frac{\epsilon}{2} + 2n \log T + 2 \inf_{\lambda \geq 0} \left(\lambda T + \frac{\sum_{t=1}^{T} \|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)} \right) \tag{4.8}
$$

The second way to bound regret is below. Observe that:

$$
\begin{aligned}
M_t(\eta_{t-1}) &= \nabla f_t(w_t)^\top (w_t - w_t^\star) - \mathrm{B}_{\tilde{g}_t}(w_t^\star \| w_t) - \frac{\epsilon}{2} \|w_t^\star - w_t\| - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(w_t^\star \| w_t) \\
&\leq \nabla f_t(w_t)^\top (w_t - w_t^\star) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(w_t^\star \| w_t) \\
&= \frac{1}{\eta_{t-1}} \left[\eta_{t-1} \nabla f_t(w_t)^\top (w_t - w_t^\star) - \mathrm{B}_{F_\psi}(w_t^\star \| w_t) \right. \\
&= \frac{1}{\eta_{t-1}} \left[\eta_{t-1} (\nabla f_t(w_t) + c\mathbf{1})^\top (w_t - w_t^\star) - \mathrm{B}_{F_\psi}(w_t^\star \| w_t) \right]
\end{aligned}
$$

Here c can be any arbitrary constant.

Using Lemma 2.10, we have the following bound if $\|\eta_{t-1}(\nabla f_t(w_t) + c\mathbf{1})\|^2_{\nabla^2 F_\psi^{-1}(w_t)} \leq \frac{1}{4}$.

$$M_t(\eta_{t-1}) \leq \frac{1}{\eta_{t-1}}\|\eta_{t-1}(\nabla f_t(w_t) + c\mathbf{1})\|^2_{\nabla^2 F_\psi^{-1}(w_t)} = \eta_{t-1}\|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2$$

Since c, can be an arbitrary constant, we have that if $\eta_{t-1}^2 \inf_c \|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2 \leq 1/4$ then:

$$M_t(\eta_{t-1}) \leq \inf_c \eta_{t-1}\|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2$$

Choosing $c = 0$, we have that if $\eta_{t-1}^2 \inf_c \|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2 \leq \eta_{t-1}^2\|\nabla f_t(w_t) \circ w_t\|_2^2 \leq 1/4$, then:

$$M_t(\eta_{t-1}) \leq \inf_c \eta_{t-1}\|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2 \leq \eta_{t-1}\|\nabla f_t(w_t) \circ w_t\|_2^2 \leq \eta_{t-1}$$

Observe that $\eta_{t-1}^2\|\nabla f_t(w_t) \circ w_t\|_2^2 \leq \frac{1}{4}$ holds if $\eta_{t-1} \leq \frac{1}{2}$. Thus, when $\eta_{t-1} \leq \frac{1}{2}$, we have $M_t(\eta_{t-1}) \leq \eta_{t-1} \leq \frac{1}{2}$ and $M_t(\eta_{t-1})/\eta_{t-1} \leq \inf_c \|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2 \leq \|\nabla f_t(w_t) \circ w_t\|_2^2 \leq 1$.

Applying Lemma 2.6, we have the following bound::

$$\begin{aligned}
\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^\gamma) &\leq \frac{\epsilon}{2} + \frac{n}{\eta_T}\log\left(\frac{1}{\gamma}\right) + \sum_{t=1}^{T} M_t(\eta_{t-1}) \\
&\leq \frac{\epsilon}{2} + n\log(1/\gamma)\left(\frac{\beta}{\alpha} + \frac{1}{2\alpha}\right) \\
&\quad + \frac{1}{2} + \sqrt{2\sum_{t=1}^{T} \inf_c \|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2}\left(\frac{n\log(1/\gamma)}{\sqrt{\alpha}} + \sqrt{\alpha}\right)
\end{aligned}$$

Using $\alpha = n\log T$, $\beta = 2n\log T$ and $\gamma = 1/T$, the above bound yields $\sum_{t=1}^{T} \tilde{f}_t(w_t) - \tilde{f}_t(w^\gamma) \leq$

$$\frac{\epsilon}{2} + \left(2n\log(T) + \frac{1}{2}\right) + \frac{1}{2} + 2\sqrt{2n\left(\sum_{t=1}^{T} \inf_c \|(\nabla f_t(w_t) + c\mathbf{1}) \circ w_t\|_2^2\right)\log(T)} \tag{4.9}$$

Both Equation (4.8) and Equation (4.9) hold simultaneously. Combining them, we have the

bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + \frac{\epsilon}{2} + 2n\log T + 2\min\left[\inf_{\lambda\geq 0}\left(\lambda T + \frac{nR}{2}\log\left(1 + \frac{\sum_{t=1}^{T}\|\hat{r}_t\|_2^2}{n(\epsilon+\lambda)}\right)\right), 1 + \sqrt{2n\left(\sum_{t=1}^{T}\inf_{c}\|(\nabla f_t(w_t) + c\mathbf{1})\circ w_t\|_2^2\right)\log(T)}\right]$$

Using $\epsilon = 1$ and $\lambda = 0$, The regret of LB-AdaCurv ONS will be:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{5}{2} + 2n\log T + \min\left(nR\log\left(1 + RT\right), 2 + 2\sqrt{2nT\log(T)}\right)$$

Using $\epsilon = 0$ and $\lambda = R/T$, the regret of LB-AdaCurv FTAL algorithm:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + 2n\log(T) + \min\left(2R + nR\log\left(1 + T^2\right), 2 + 2\sqrt{2nT\log(T)}\right)$$

■

LB-AdaCurv ONS/FTAL maintains the $O(nR\log T)$ data-dependent regret of AdaCurv ONS/FTAL while guaranteeing a worst-case regret of $O(\sqrt{nT\log T})$.

4.5 More Data-Dependent Regret Bounds

In this section, we study two more kinds of data-dependent regret bounds.

4.5.1 First-Order Regret Bound

Let $L_T^\star = \min_{w\in\Delta_n}\left[\sum_{t=1}^{T} f_t(w)\right] - \sum_{t=1}^{T}\left[\min_{w\in\Delta_n} f_t(w)\right]$ be the regret between the best static and the best dynamic portfolio selection strategies. In the worst case, $L_T^\star = O(T)$. However, if the returns are not adversarial generated, then $L_T^\star$ could be quite small. Bounds which depend on $L_T^\star$ instead of T are termed First-Order bounds.

Orabona *et al.* [66] show that ONS has the regret bound $O((C/c)n \log L_T^\star)$, when the returns are in $[c, C]^n$. Using the recent UOCO algorithm from Yan *et al.* [67], we can obtain a regret of $O(Rn \log L_T^\star)$ for any sequence of returns. However, these do not guarantee a bounded worst-case regret. Tsai *et al.* [68] were able to obtain a bound of $O(\sqrt{dL_T^\star} \log T)$ for any sequence of returns, with a run time of $O(n)$. As $L_T^\star = O(T)$, it guarantees a bounded worst-case regret. We show that the AdaCurv ONS algorithm achieves $O(Rn \log L_T^\star)$ regret. Moreover, by adding extra regularization via the log-barrier, the LB-AdaCurv ONS algorithm achieves a regret of $O(\min(Rn \log(L_T^\star), \sqrt{nL_T^\star \log T}))$. These results are summarized in Table 4.3.

Table 4.3: First-Order Regret Bounds for Online Portfolio Selection

Algorithm	**First-Order Regret**	**Worst-case Regret**	**Run-time**
UOCO Yan *et al.* [67]	$Rn \log L_T^\star$	∞	$n^3 \log T$
Tsai *et al.* [68]	$\sqrt{nL_T^\star} \log T$	$\sqrt{nT} \log T$	n
AdaCurv ONS	$Rn \log(L_T^\star)$	∞	n^3
LB-AdaCurv ONS	$\min(Rn \log(L_T^\star), \sqrt{nL_T^\star \log T})$	$\sqrt{nT \log T}$	n^3

Theorem 4.4. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, define $f_t(w) = -\log(r_t^\top w)$. The updates of AdaCurv ONS (Equation (4.5)) with $\epsilon = 1$ satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{1}{2} + \frac{nR}{2} \log\left(4nR^3 \log\left(\frac{4nR^3}{e}\right) + 4R^2 + 8R^2 L_T^\star + 2\right)$$

Here, $L_T^\star = \min_{w \in \Delta_n} \left[\sum_{t=1}^T f_t(w)\right] - \sum_{t=1}^T \left[\min_{w \in \Delta_n} f_t(w)\right]$ is the regret between the best static and the best dynamic portfolio selection strategies.

Theorem 4.5. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, define $f_t(w) = -\log(r_t^\top w)$. The updates of LB-AdaCurv ONS (Algorithm 2) with $\epsilon = 1$ satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w_t) \leq \frac{5}{2} + 2n \log T + \min\left[2 + 2\sqrt{8n \log T} + 4\sqrt{8n\left(L_T^\star + \frac{9}{2} + 2n \log T\right) \log T},\right.$$

$$nR \log \left(8nR^3 \log \left(\frac{8nR^3}{e} \right) + 20R^2 + 16R^2 n \log T + 8R^2 L_T^\star + 2 \right) \Bigg]$$

Here, $L_T^\star = \min_{w \in \Delta_n} \left[\sum_{t=1}^T f_t(w) \right] - \sum_{t=1}^T \left[\min_{w \in \Delta_n} f_t(w) \right]$ *is the regret between the best static and the best dynamic portfolio selection strategies.*

The proofs of the above theorems appear in Appendix B.

4.5.2 Second-Order Regret Bound

Hazan and Kale [62] show that the Exp-Concave FTL algorithm has a regret of $O(R^2 n \log Q_T)$, where $Q_T = \min_x \sum_{t=1}^T \|r_t - x\|_2^2$ and a per-iteration run time of $O(n^{2.5}T)$. They also proposed an algorithm that uses a quadratic surrogate, called Faster Quadratic-variation Universal Algorithm (FQUA). This has a regret bound of $O((C/c)^3 n \log Q_T)$ and a run-time of $O(n^3)$ per round when the returns are in $[c, C]^n$. Recently, Tsai *et al.* [68], showed a second order bound $O\left(\sqrt{n\tilde{Q}_T \log T}\right)$, where $\tilde{Q}_T = \min_x \sum_{t=1}^T \left\| \frac{r_t \circ w_t}{r_t^\top w_t} - x \right\|_2^2$ with a run time of $O(n \log \log T)$ for any sequence of returns. Since $\tilde{Q}_T = O(T)$, their algorithm has a worst-case regret bound of $O(\sqrt{nT \log T})$. However, the definition of $\tilde{Q}_T$ has no meaningful interpretation in terms of quadratic variation of returns r_t.

We show that modifying FQUA with adaptive-curvature surrogate functions has regret $O(Rn \log Q_T)$ with a run time of $O(n^3)$. This algorithms is termed AdaCurv ONS. These prior works, along with our contributions are summarized in Table 4.4.

Table 4.4: Second Order Regret Bounds for Online Portfolio Selection

Algorithm	**Second Order Regret**	**Worst-case Regret**	**Run-time (per round)**
Exp-Concave FTL [62]	$R^2 n \log(Q_T)$	∞	$n^{2.5}T$
Tsai *et al.* [68]	$\sqrt{n\tilde{Q}_T \log T}$	$\sqrt{nT \log T}$	$n \log \log T$
AdaCurv ONS	$R^2 n \log(Q_T)$	∞	n^3

Theorem 4.6. *For* $w \in \Delta$*, any sequence of returns* $r_1, \ldots, r_T \in \mathbb{R}_+^n$*, define* $f_t(w) = -\log(r_t^\top w)$*.*

The AdaCurv ONS updates (Equation (4.5)*) with* $\epsilon = 1$ *satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) = O\left(nR^2 \log(1 + Q_T + n) + \sqrt{n}R \log(1 + Q_T/n) + 1\right)$$

Here $Q_T = \min_\mu \sum_{t=1}^{T} \|r_t - \mu\|_2^2 = \sum_{t=1}^{T} \|r_t - \bar{r}_T\|_2^2$*, where* $\bar{r}_T = \frac{1}{T}\sum_{t=1}^{T} r_t$*.*

The proof of the above theorem appears in Appendix B.

4.6 Conclusion

In this chapter, we studied data-dependent regret bounds for the Online Portfolio Selection problem. We first obtained a new adaptive curvature surrogate function for the loss $-\log(r_t^\top w)$. We presented two algorithms that use this surrogate function. First is the AdaCurv ONS algorithm that has a data-dependent regret bound of $O(nR \log T)$, where $R = \max_{t,i,j} r_t(i)/r_t(j)$. Next, we proposed the LB-AdaCurv ONS algorithm, which adds the log-barrier regularization to AdaCurv ONS, giving it a regret of $O(\min(nR \log T, \sqrt{nT \log T}))$. In addition, we showed that AdaCurv ONS has a logarithmic first-order regret bound of $O(nR \log L_T^\star)$ and LB-AdaCurv ONS has a first-order regret bound of $O(\min(nR \log L_T^\star, \sqrt{nL_T^\star \log T}))$. Finally, AdaCurv ONS is also shown to obtain a second-order regret bound of $O(nR^2 \log Q_T)$ with a run time of $O(n^3)$, improving over the run time of the Exp-Concave FTL algorithm of Hazan and Kale [62]. Finding an algorithm achieving a logarithmic second order regret bound in Q_T, while also having a bounded worst case regret remains open problem.

The ONS algorithm [3] with the constant curvature surrogate functions is an important building block in recent OPS algorithms such as AdaBARRONS [57], BISONS [58] and PAE+DONS [59]. As future work, one could explore if the usage of our adaptive curvature surrogate function along with the techniques developed in [57, 58, 59] could lead to more elegant algorithms as it avoids the need to tune β in the quadratic surrogate.

In the next chapter, we study yet another kind of data-dependent regret bound called the gradual-variation bound for OPS.

Chapter 5: Online Portfolio Selection with Predicted Returns

5.1 Introduction

In the classical version of the online portfolio selection (OPS) problem, it is assumed that the investor does not have any prior belief about the future returns. At time t, the investor only uses the returns $r_1, \ldots, r_{t-1}$ and possibly prior portfolios $w_1, \ldots, w_{t-1}$ to select the portfolio vector w_t. However, in practice, investors do have prior beliefs about future returns. This prior belief is typically expressed as a distribution D_t over the future returns of assets, i.e., the investor believes that r_t would be a random variable drawn from D_t. Thus the distribution D_t is the investor's prediction for the current return. The actual market return r_t however, could be quite different from the investor's prediction. In this chapter we study algorithms that incorporate the investor's predictions to further reduce regret. As such, the specific process by which an investor creates these predictions in each round is not the subject of our study. Instead, we study how an investor could incorporate such predictions into the online portfolio selection framework. The interaction protocol for our setting is stated below:

Online Portfolio Selection with Predictions - Interaction Protocol:

Initial information set $\mathcal{I}_1 = \{\}$

for $t = 1$ *to* T **do**

- Investor receives prediction D_t distribution. Augmented information set $\mathcal{I}_t \cup \{D_t\}$
- Investor picks the portfolio $w_t \in \Delta_n$
- Market revels $r_t \in \mathbb{R}^n$
- Investor's wealth grows by a multiplicative factor of $r_t^\top w_t$
- Update information set $\mathcal{I}_{t+1} = I_t \cup \{w_t, r_t\}$

The investor could leverage their predictions to pick a portfolio in several possible ways. In his seminal work, which birthed the field of *Modern Portfolio Theory*, Markowitz [69] proposed using

an optimization framework that balances the expected return of a portfolio with its variance. Given a predicted returns distribution D_t, Markowitz's approach of mean-variance optimization can be stated as:

$$w_t \in \arg\max_{w \in \Delta_n} \mathbb{E}_{r \sim D_t}[r^\top w] - \frac{\lambda}{2} \mathbb{V}ar_{r \sim D_t}[r^\top w]$$

Here λ is a parameter capturing the investor's aversion to the variance of the returns according to the investor's predictions. While Markowitz's approach is widely used in practice, it's actual performance relies on the accuracy of the investor's predictions. In Markowitz's framework the investor is said to be maximizing risk-adjusted return.

Capital Growth Theory, developed by Kelly [70] and Thorp [71], proposes a different solution for an investor who seeks to maximize the growth rate of their wealth. This approach is more suitable for a multi-period portfolio selection problem like ours. Given a predicted returns distribution D_t, the Kelly criteria proposes picking the log-optimal portfolio:

$$w_t \in \arg\max_{w \in \Delta_n} \mathbb{E}_{r \sim D_t}[\log(r^\top w)]$$

The performance of the log-optimal portfolio is similarly sensitive to the accuracy of the investor's predictions.

Both of Markowitz's and Kelly's approaches are special cases of *Expected Utility Theory*, which studies decision making under uncertainty. Given an investor with a specified concave utility function U and return prediction D_t, the investor picks w_t as:

$$w_t \in \arg\max_{w \in \Delta_n} \mathbb{E}_{r \sim D_t}[U(r^\top w)]$$

The performance of an expected utility investor is only as good as the accuracy of their prediction. Thus, in practice most active investors expend a tremendous amount of effort in coming up with accurate predictions. Moreover, unlike OPS techniques stemming from *Regret Minimization Theory* (like the ones discussed in Chapter 4), expected utility theory does not usually provide

any kind of worst-case or data-dependent performance guarantee for the investor. On the other hand, as expected utility theory uses predictions to make decisions, an expected utility investor will outperform a purely regret minimizing investor if the predictions are indeed accurate.

In this chapter, we hope to combine the two approaches of expected utility and regret minimization, while maintaining their advantages. We initiate the study of online portfolio selection algorithms with predicted returns. At time t, these algorithms not only use the returns $r_1, \ldots, r_{t-1}$ and prior portfolios $w_1, \ldots, w_{t-1}$, but also use any available return predictions D_t to select the portfolio vector w_t. Online decision making algorithms that incorporate predictions have been studied under the name of *Online Learning with Predictions* (OLP) [12] in the online optimization literature, where the goal is to improve regret. They have also been studied under the name of *Algorithms with Predictions* [25] in the online algorithms literature, where the goal is to improve competitive ratio. Since our study of the OPS problem uses techniques from OCO, we focus on the online optimization approach.

In both these areas, algorithms that incorporate predictions seek to achieve some form of *consistency* and *robustness* guarantees. Consistency implies that the algorithm should be able to improve its performance by taking advantage of the predictions in case they are accurate. Robustness implies that the algorithm should retain its worst-case performance guarantee in case the predictions have large errors or are misspecified. In the context of OPS, we present two algorithms that achieve different consistency and robustness guarantees. We can think of our algorithms as strategies for a hybrid investor that aims to achieve a trade-off between an expected utility investor and regret minimizing investor.

In this chapter, we first present an algorithm based on Optimistic FTRL algorithm that achieves a worst case static-regret of $O(n \log T)$ when the predictions are exact and $O(\sqrt{nT \log T})$ when the predictions are completely arbitrary. In other words, we say that this algorithm is $O(n \log T)$-consistent and $O(\sqrt{nT \log T})$-robust with respect to static regret. Recall that worst-case static-regret bounds the quantity $\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w^\star)$, where $w^\star \in \arg\min_{w \in \Delta_n} \sum_{t=1}^{T} f_t(w)$ is the best static allocation. In the presence of predictions, it is important to also consider the regret

with respect to the expected utility investor, i.e., the quantity $\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w_t^{EU})$, where $w_t^{EU} \in \arg\max_{w \in \Delta_n} \mathbb{E}_{r \sim D_t}[U(r^\top w)]$. To bound this, we propose a second algorithm that combines the portfolios of an expected utility investor and a regret minimizing investor to achieve simultaneously a $O(\log T)$ regret against the expected utility player and $O(n \log T)$ regret against the best performing static allocation $w^\star$.

5.2 Optimistic Expected Utility LB-FTRL

Consider an investor who at time t has a prediction distribution D_t and a utility function U. We can augment the log-barrier regularized FTRL algorithm with the expected utility of the player by using the following update:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} \nabla f_t(w_t)^\top w + \frac{F_\psi(w)}{\eta_{t-1}} - \mathbb{E}_{r \sim D_t}[U(r^\top w)]$$

Here $F_\psi(w) = \sum_{i=1}^{n} [\log(1/n) - \log(w(i))]$ is the log-barrier regularizer. The parameter η_t is tuned via the AdaFTRL technique as:

$$\eta_t = \frac{n \log T}{Cn \log T + \sum_{s=1}^{t} M_s(\eta_{s-1})}$$

$M_t(\eta_{t-1})$ is defined as:

$$M_t(\eta_{t-1}) = \sup_{w \in \Delta_n} \left(\nabla f_t(w_t) + \mathbb{E}_{r \sim D_t}[U'(r^\top w_t) r]\right)^\top (w_t - w) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(w \| w_t)$$

Here U' is the first derivative of U and C is a constant chosen such that $C = 1 + \sup_x xU'(x)$.

The algorithm is also summarized below:

Theorem 5.1. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, return prediction distributions $D_1, \ldots, D_T$, concave and strictly increasing utility function U with a strictly decreasing first derivative U', define $f_t(w) = -\log(r_t^\top w)$. The updates of OEU-LB-FTRL (Algorithm 3) satisfy the*

Algorithm 3: Optimistic Expected Utility Log-Barrier FTRL (OEU-LB-FTRL)

Starting Parameters: $\eta_0 = 1/2$

Pick $C = 1 + \sup_x xU'(x)$

Regularizer $F_\psi(q) = \sum_{i=1}^{n} (f_\psi(w(i)) - f(1/n))$, where $f_\psi(x) = -\log(x)$

for $t = 1$ *to* T **do**

Investor has a prediction distribution D_t and utility function U_t

Pick portfolio:

$$w_t \in \arg\min_{w \in \Delta_n} \sum_{s=1}^{t-1} \nabla f_t(w_t)^\top w + \frac{F_\psi(w)}{\eta_{t-1}} - \mathbb{E}_{r \sim D_t}[U(r^\top w)]$$

Observe returns vector r_t. Let $f_t(w) = -\log(r_t^\top w)$

Let $\tilde{g}_t = \sum_{s=1}^{t} \tilde{f}_t$. Compute $M_t(\eta_{t-1})$ by solving the optimization:

$$M_t(\eta_{t-1}) = \sup_{w \in \Delta_n} \left[\left(\nabla f_t(w_t) + \mathbb{E}_{r \sim D_t}[U'(r^\top w_t) r] \right)^\top (w_t - w) - \frac{1}{\eta} \mathrm{B}_{F_\psi}(w \| w_t) \right]$$

Compute η_t:

$$\eta_t = \frac{n \log T}{Cn \log T + \sum_{s=1}^{t} M_s(\eta_{s-1})}$$

regret bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + C\,(1 + 2n\log T) + 2\sqrt{2n\left(\sum_{t=1}^{T}\left\|\mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t}\right\|_2^2\right)\log T}$$

Where $C = 1 + \sup_x xU'(x)$. *This implies the worst-case regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + C\,(1 + 2n\log T) + 2C\sqrt{2nT\log T}$$

Moreover, if U *and* D_t *are such that* $\mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] = \frac{r_t \circ w_t}{r_t^\top w_t}$, *then we have the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + 2Cn\log T$$

Proof. First, we decompose the regret into two terms:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) = \sum_{t=1}^{T} f_t(w_t) - f_t(w^\gamma) + \sum_{t=1}^{T} f_t(w^\gamma) - f_t(w)$$

Here $w^\gamma = (1-\gamma)w + \gamma/n$. For the second term, we have:

$$\sum_{t=1}^{T} f_t(w^\gamma) - f_t(w) = \sum_{t=1}^{T} \log\left(\frac{r_t^\top w}{(1-\gamma) r_t^\top w + \gamma r_t^\top \frac{1}{n}}\right) \leq T\log\left(\frac{1}{1-\gamma}\right) \leq 2\gamma T$$

Here, we used the fact that when $\gamma \leq 1/2$, we have $\log\left(\frac{1}{1-\gamma}\right) \leq 2\gamma$. For the first term, we use convexity:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w^\gamma) \leq \sum_{t=1}^{T} \nabla f_t(w_t)^\top (w_t - w^\gamma)$$

So, we have:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \sum_{t=1}^{T} \nabla f_t(w_t)^\top (w_t - w^\gamma) + 2\gamma T$$

Since OEU-LB-FTRL is an instance of optimistic FTRL on the gradients $\nabla f_t(w_t)$, we can apply Theorem 2.3 with hint function $m_t(w) = -\mathbb{E}_{r\sim D_t}[U(r^\top w)]$. This gives the regret inequality $\nabla f_t(w_t)^\top (w_t - w^\gamma)$:

$$\begin{aligned}
&\leq \frac{F_\psi(w^\gamma)}{\eta_T} + \sum_{t=1}^{T} \left[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \frac{\mathrm{B}_{F_\psi}(w'_{t+1} \| w_t)}{\eta_{t-1}} \right] \\
&\leq \frac{F_\psi(w^\gamma)}{\eta_T} + \sum_{t=1}^{T} M_t(\eta_{t-1})
\end{aligned}$$

Note that $\nabla m_t(w) = -\mathbb{E}_{r\sim D_t}[U'(r^\top w) r]$ and w'_t are the iterates of LB-FTRL with no hints. We bound $F_\psi(w^\gamma)$ below:

$$\begin{aligned}
F_\psi(w^\gamma) &= n \log(1/n) - (n-1)\log(\gamma/n) - \log((1-\gamma) + \gamma/n) \\
&\leq n \log(1/n) - n \log(\gamma/n) = n \log(1/\gamma)
\end{aligned}$$

Consider $M_t(\eta_{t-1})$. Assume the supremum in it's optimization occurs at $w_t^\star$.

$$\begin{aligned}
M_t(\eta_{t-1}) &= \left(\nabla f_t(w_t) + \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r]\right)^\top (w_t - w_t^\star) - \frac{1}{\eta_{t-1}} \mathrm{B}_{F_\psi}(w_t^\star \| w_t) \\
&= \frac{1}{\eta_{t-1}} \left(\eta_{t-1} \left(\nabla f_t(w_t) + \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r]\right)^\top (w_t - w_t^\star) - \mathrm{B}_{F_\psi}(w_t^\star \| w_t) \right)
\end{aligned}$$

Applying Lemma 2.10, if $\| \eta_{t-1} \left(\nabla f_t(w_t) + \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r]\right) \|_{\nabla^2 F_\psi(w_t)^{-1}} \leq 1/2$, then:

$$\begin{aligned}
M_t(\eta_{t-1}) &\leq \frac{1}{\eta_{t-1}} \| \eta_{t-1} \left(\nabla f_t(w_t) + \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r]\right) \|^2_{\nabla^2 F_\psi(w_t)^{-1}} \\
&= \eta_{t-1} \left\| \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 \\
&= \eta_{t-1} \left\| \mathbb{E}_{r\sim D_t} \left[(r^\top w_t) U'(r^\top w_t) \frac{r \circ w_t}{r^\top w_t} \right] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 \\
&\leq \eta_{t-1} (1 + \sup_x x U'(x))^2
\end{aligned}$$

Since $\|\nabla f_t(w_t) + \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r]\|_{\nabla^2 F_\psi(w_t)^{-1}} \leq 1 + \sup_x x U'(x)$, we can ensure that Lemma 2.10

is applicable by picking $\eta_{t-1} \leq (2(1 + \sup_x xU'(x)))^{-1}$. Thus, we have

$$M_t(\eta_{t-1}) \leq \eta_{t-1}(1 + \sup_x xU'(x))^2 \leq \frac{1 + \sup_x xU'(x)}{2} = \frac{C}{2}$$

$$\frac{M_t(\eta_{t-1})}{\eta_{t-1}} \leq \left\| \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 \leq (1 + \sup_x xU'(x))^2 = C^2$$

Applying Lemma 2.6 and picking $\gamma = 1/T$, we have the bound:

$$\frac{F_\psi(w^\gamma)}{\eta_T} + \sum_{t=1}^{T} M_t(\eta_{t-1}) \leq C\left(2n\log T + \frac{1}{2}\right) + \frac{C}{2} + 2\sqrt{2n\left(\sum_{t=1}^{T} \left\| \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2\right)\log T}$$

Thus, we have the first bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + C\,(1 + 2n\log T) + 2\sqrt{2n\left(\sum_{t=1}^{T} \left\| \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2\right)\log T}$$

Since $\left\| \mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 \leq C^2$, this implies the worst-case bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + C\,(1 + 2n\log T) + 2C\sqrt{2nT\log T}$$

If the condition $\mathbb{E}_{r\sim D_t}[U'(r^\top w_t) r \circ w_t] = \frac{r_t \circ w_t}{r_t^\top w_t}$ is satisfied, then $M_t(\eta_{t-1}) = 0$. In this case, we have the bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + \frac{F_\psi(w^\gamma)}{\eta_T} \leq 2 + 2Cn\log T$$

■

For the specific case of a Kelly Criteria investor, i.e., $U(x) = \log(x)$ the above theorem reduces to the following corollary.

Corollary 5.2. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, return prediction distribu-*

tions $D_1, \ldots, D_T$, define $f_t(w) = -\log(r_t^\top w)$. The updates of OEU-LB-FTRL (Algorithm 3) with $U(x) = \log(x)$ satisfy the regret bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 4 + 4n \log T + 2\sqrt{2n \left(\sum_{t=1}^{T} \left\| \mathbb{E}_{r \sim D_t} \left[\frac{r \circ w_t}{r^\top w_t} \right] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 \right) \log T}$$

This implies the worst-case regret bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 4 + 4n \log T + 4\sqrt{nT \log T}$$

Moreover, if D_t is such that $\mathbb{E}_{r \sim D_t} \left[\frac{r \circ w_t}{r^\top w_t} \right] = \frac{r_t \circ w_t}{r_t^\top w_t}$, then we have the regret bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 2 + 4n \log T$$

Proof. For $U(x) = \log(x)$, the constant $C = 1 + \sup_x xU'(x) = 2$. Moreover, we have the bound:

$$\left\| \mathbb{E}_{r \sim D_t} [U'(r^\top w_t) r \circ w_t] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 = \left\| \mathbb{E}_{r \sim D_t} \left[\frac{r \circ w_t}{r^\top w_t} \right] - \frac{r_t \circ w_t}{r_t^\top w_t} \right\|_2^2 \leq 2$$

■

5.2.1 Robustness and Consistency

The two regret bounds in Corollary5.2 showcase the robustness and consistency properties of Algorithm 3. As the first regret bound of $O(\sqrt{nT \log T})$ holds for any prediction distribution D_t, it constitutes a robustness guarantee. In the scenario where D_t is a delta distribution on r_t, we have $\mathbb{E}_{r \sim D_t} \left[\frac{r \circ w_t}{r^\top w_t} \right] = \frac{r_t \circ w_t}{r_t^\top w_t}$, so the regret is $O(n \log T)$. Thus, Algorithm 3 with $U(x) = \log(x)$ is $O(n \log T)$-consistent and $O(\sqrt{nT \log T})$-robust with respect to static regret.

When viewed from the perspective of wealth generated, the robustness consistency guarantee of Theorem 5.1 is somewhat unsatisfactory. Denote by $W(Alg)$ the wealth of the investor and $W(w^\star)$ the wealth of the best static allocation vector $w^\star \in \arg\min_{w \in \Delta_n} \sum_{t=1}^{T} f_t(w)$. Then, the robustness

guarantee ensures that $W(Alg) \geq W(w^\star)\exp(-O(\sqrt{nT\log T}))$. If the prediction distribution D_t perfectly predicts r_t, then the consistency guarantee ensures $W(Alg) \geq W(w^\star)\exp(-O(n\log T))$. Considering the fact that a purely regret minimizing investor who uses Cover's algorithm can ensure $W(Cover) \geq W(w^\star)\exp(-O(n\log T))$, without even taking the predictions D_t into account, we can see that the consistency guarantee of $W(Alg) \geq W(w^\star)\exp(-O(n\log T))$ with perfect predictions is quite weak. Additionally, with perfect predictions, the wealth of the expected utility player $W(EU)$ will be $\prod_{t=1}^{T}(\max_i r_t(i))$, which could be exponentially larger than the wealth of the best static allocation $W(w^\star) = \prod_{t=1}^{T}(r_t^\top w^\star)$.

Ideally, we would like to seek an algorithm whose robustness guarantee ensures that the investor's wealth is close to $W(w^\star)$ in case the prediction distributions D_t are completely arbitrary. Thus, it should at least perform like a purely regret minimizing algorithm. Additionally, if the predictions are perfect, then the consistency guarantee must ensure that investor's wealth is close to $W(EU)$. In the next section, we present an algorithm that achieves this robustness-consistency guarantee. As the wealth achieved will always track the better of $W(EU)$ or $W(w^\star)$ depending on the accuracy of predictions, our algorithm obtains the best of both worlds.

5.3 Best of Both Worlds for Online Portfolio Selection

Consider an expected utility investor (EU) who picks a portfolio w_t^{EU} at time t by solving the stochastic optimization problem: $w_t^{EU} = \arg\max_{w\in\Delta_n} \mathbb{E}_{r\sim D_t}\left[U(r^\top w)\right]$. Also consider a purely regret minimizing investor RM) who picks a portfolio w_t^{RM} in round t by using a regret minimizing algorithm like the ones in Table 4.1. Denote the algorithm used by this player as RM. It takes as input the information set $\{w_1^{RM}, r_1, \ldots, w_{t-1}^{RM}, r_{t-1}\}$ and outputs the allocation w_t^{RM}. Finally consider an meta-investor who can only allocates wealth to EU or RM. The meta-investor cannot directly participate in the actual market of n assets, but can indirectly participate via the two base investors, EU and RM. At time t, the meta-investor allocates γ_t portion of wealth to EU and $1-\gamma_t$ portion to RM. The quantity γ_t itself could be chosen by using a regret minimizing algorithm from Table 4.1. Denote the algorithm used by the meta-investor as $metaRM$. The returns seen by meta-

investor are the returns of the base investors, $r_t^{EU} = r_t^\top w_t^{EU}$ and $r_t^{RM} = r_t^\top w_t^{RM}$. So, if the allocation chosen by $metaRM$ is γ_t, then the implicit allocation of the meta-investor is $\gamma_t w_t^{EU} + (1-\gamma_t) w_t^{RM}$. The Best of Both Worlds OPS algorithm picks $w_t = \gamma_t w_t^{EU} + (1-\gamma_t) w_t^{RM}$.

Algorithm 4: Best of Both Worlds for Online Portfolio Selection (BoB-OPS)

for $t = 1$ *to* T **do**

- EU investor picks portfolio: $w_t^{EU} = \arg\max_{w \in \Delta_n} \mathbb{E}_{r \sim D_t} \left[U(r^\top w) \right]$
- RM investor picks portfolio: $w_t^{RM} = RM(\{w_1^{RM}, r_1, \ldots, w_{t-1}^{RM}, r_{t-1}\})$
- Pick EU-RM allocation $\gamma_t = metaRM(\{\gamma_1, (r_1^{EU}, r_1^{RM}), \ldots, \gamma_{t-1}, (r_{t-1}^{EU}, r_{t-1}^{RM})\})$
- Pick portfolio: $w_t = \gamma_t w_t^{EU} + (1-\gamma_t) w_t^{RM}$
- See returns r_t. Set $r_t^{EU} = r_t^\top w_t^{EU}$ and $r_t^{RM} = r_t^\top w_t^{RM}$

Theorem 5.3. *For $w \in \Delta_n$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, return prediction distributions $D_1, \ldots, D_T$, concave and strictly increasing utility function U with a strictly decreasing first derivative U', define $f_t(w) = -\log(r_t^\top w)$. The updates of BoB-OPS (Algorithm 4) satisfy the regret bounds:*

$$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w_t^{EU}) \leq \mathcal{R}_{metaRM}(2, T)$$

$$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w) \leq \mathcal{R}_{RM}(n, T) + \mathcal{R}_{metaRM}(2, T)$$

$\mathcal{R}_{RM}(n, T)$ and $\mathcal{R}_{metaRM}(2, T)$ are the regret bounds for the algorithm used by the RM investor and the meta-investor respectively.

Proof. The regret bound for the meta-investor is:

$$\sum_{t=1}^{T} -\log(\gamma_t r_t^{EU} + (1-\gamma_t) r_t^{RM}) - \sum_{t=1}^{T} -\log(\gamma r_t^{EU} + (1-\gamma) r_t^{RM}) \leq \mathcal{R}_{metaRM}(2, T)$$

Pick $\gamma = 1$ and note that $r_t^{EU} = r_t^\top w_t^{EU}$, $r_t^{RM} = r_t^\top w_t^{RM}$

$$\implies \sum_{t=1}^{T} -\log(\gamma_t r_t^\top w_t^{EU} + (1-\gamma_t) r_t^\top w_t^{RM}) - \sum_{t=1}^{T} -\log(r_t^\top w_t^{EU}) \leq \mathcal{R}_{metaRM}(2, T)$$

$$\implies \sum_{t=1}^{T} -\log(r_t^\top w_t) - \sum_{t=1}^{T} -\log(r_t^\top w_t^{EU}) \leq \mathcal{R}_{metaRM}(2,T)$$
$$\implies \sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w_t^{EU}) \leq \mathcal{R}_{metaRM}(2,T)$$

This gives the first bound in the theorem. If we pick $\gamma = 0$, we would have arrived at:

$$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w_t^{RM}) \leq \mathcal{R}_{metaRM}(2,T)$$

The regret bound for the RM investor is:

$$\sum_{t=1}^{T} f_t(w_t^{RM}) - \sum_{t=1}^{T} f_t(w) \leq \mathcal{R}_{RM}(n,T)$$

Adding these two inequalities, we have the second bound in the theorem:

$$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w) \leq \mathcal{R}_{metaRM}(2,T) + \mathcal{R}_{RM}(n,T)$$

■

We show a concrete bound by instantiating RM and $metaRM$ with Cover's Universal Portfolio algorithm. Regret bound for Cover's UP is $\mathcal{R}_{UP}(n,T) = (n-1)\log(T+1)$ [4, Theorem 10.3].

Corollary 5.4. *For $w \in \Delta_n$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, return prediction distributions $D_1, \ldots, D_T$, concave and strictly increasing utility function U with a strictly decreasing first derivative U', define $f_t(w) = -\log(r_t^\top w)$. The updates of BoB-OPS (Algorithm 4) where the RM invetsor and meta-RM investor use Cover's Universal Portfolio[1] algorithm satisfy the regret bounds:*

$$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w_t^{EU}) \leq \log(T+1)$$

$$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w) \leq n\log(T+1)$$

Corollary 5.4 implies the following wealth lower bound:

$$W(\text{Bob-OPS}) \geq \max\left(\frac{W(EU)}{T+1}, \frac{W(w^\star)}{(T+1)^n}\right)$$

Thus, when the predictions D_t are perfect, we have the consistency guarantee $W(\text{Bob-OPS}) \geq \frac{W(EU)}{T+1}$. When the predictions are arbitrarily bad, we have the robustness guarantee $W(\text{Bob-OPS}) \geq \frac{W(w^\star)}{(T+1)^n}$, where $w^\star \in \arg\min_{w\in\Delta_n} \sum_{t=1}^T f_t(w)$ is the optimal static allocation in hindsight.

5.4 Gradual-Variation Bound

Gradual-Variation Bounds are data dependent regret bounds where the regret is bounded as a measure of variation between consecutive returns. For the OPS problem, these bounds were first studied in Chiang *et al.* [21], who showed that an optimistic variant of the ONS algorithm has a regret of $O(nC^2 \log V_T)$ when the returns are in $[c, C]^n$. Here $V_T = \sum_{t=1}^T \|r_t - r_{t-1}\|_2^2$. The UOCO algorithm of [67] obtains a gradual variation bound of the form $O(nR \log V_T)$. Due to the dependence on R the worst-case regret of this approach is not bounded.

Tsai *et al.* [68] obtain a regret bound of $O(\sqrt{n\tilde{V}_T} \log T)$, where $\tilde{V}_T = \sum_{t=2}^T \left\| \frac{r_t \circ w_{t-1}}{r_t^\top w_{t-1}} - \frac{r_{t-1} \circ w_{t-1}}{r_t^\top w_{t-1}} \right\|_2^2$. It implies a worst case $O(\sqrt{nT} \log T)$ regret bound. Their algorithm is an instance of log-barrier FTRL and uses *multiplicative-gradient optimism*, which is an implicit technique for simultaneously guessing the gradient $\nabla f_t(w_t)$ and picking the portfolio w_t. Their algorithm does not allow for specifying a utility function or predicted distribution like our Algorithm 3.

We are able to obtain a similar gradual-variation bound as Tsai *et al.* [68] by setting the current prediction D_t as a delta distribution on r_{t-1} and using the logarithmic utility function. The update equation in this case can be stated as:

$$w_t \in \arg\min_{w\in\Delta_n} \sum_{s=1}^{t-1} \nabla f_t(w_t)^\top w + f_{t-1}(w) + \frac{F_\psi(w)}{\eta_{t-1}}$$

Thus, it can be analyzed as a LB-FTRL with hint function $m_t = f_{t-1}$.

Corollary 5.5. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}^n_+$, let the return prediction distribution D_t be the delta distribution on r_{t-1} (Let r_0 be the all 1s vector). The updates of OEU-LB-FTRL (Algorithm 3) with $U(x) = \log(x)$ satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq 4 + 4n \log T + 2\sqrt{2n\tilde{V}'_T \log T}$$

Here $\tilde{V}'_T = \sum_{t=1}^{T} \left\| \frac{r_t \circ w_t}{r_t^\top w_t} - \frac{r_{t-1} \circ w_t}{r_{t-1}^\top w_t} \right\|_2^2$ and r_0 is the all ones vector.

Proof. Apply Corollary 5.2 with D_t being the delta distribution on r_{t-1}. ■

Since $\tilde{V}'_T = \sum_{t=1}^{T} \left\| \frac{r_t \circ w_t}{r_t^\top w_t} - \frac{r_{t-1} \circ w_t}{r_{t-1}^\top w_t} \right\|_2^2 \leq \sum_{t=1}^{T} \sup_{w \in \Delta_n} \left\| \frac{r_t \circ w}{r_t^\top w} - \frac{r_{t-1} \circ w}{r_{t-1}^\top w} \right\|_2^2 \leq 2T$, we have a $O(\sqrt{nT \log T})$ worst-case regret bound. Table 5.1 summarizes the results on gradual-varioation bounds.

Table 5.1: Gradual-Variation Regret Bounds for Online Portfolio Selection

Algorithm	**First-Order Regret**	**Worst-case Regret**	**Run-time**
UOCO Yan *et al.* [67]	$Rn \log V_T$	∞	$n^3 \log T$
Tsai *et al.* [68]	$\sqrt{n\tilde{V}_T} \log T$	$\sqrt{nT} \log T$	n
OEU-LB-FTRL (Corollary 5.5)	$\sqrt{n\tilde{V}'_T \log T}$	$\sqrt{nT \log T}$	n^3

5.5 Conclusion

In this chapter, we have explored the integration of predicted returns into the online portfolio selection (OPS) framework, presenting a novel approach that bridges the gap between regret minimization techniques and expected utility theory to guide investment decisions. By incorporating the investor's prior beliefs about asset returns through distribution D_t, we introduced algorithms that aim to optimize portfolio selection by balancing the inherent trade-off between exploiting these predictions to improve performance and maintaining robustness against prediction inaccuracies.

We presented two main algorithmic contributions: the Optimistic Expected Utility Log-Barrier Follow-the-Regularized-Leader (OEU-LB-FTRL) algorithm and the Best of Both Worlds for On-

line Portfolio Selection (BoB-OPS) algorithm. The OEU-LB-FTRL algorithm demonstrates how investors can adjust their portfolios by considering both past returns and future return predictions. This algorithm is shown to be both consistent and robust with respect to static regret. The OEU-LB-FTRL algorithm with logarithmic utility is shown to be $O(n \log T)$-consistent and $O(\sqrt{nT \log T})$-robust with respect to static regret.

The BoB-OPS algorithm allocates wealth between a purely regret minimizing investor (RM) and an expected utility investor (EU) based on their past performance. This approach not only allows investors to benefit from accurate predictions when available but also ensures that their performance does not drastically suffer from poor predictions. The theoretical analysis confirms that BoB-OPS can achieve the best of both worlds, adapting to the accuracy of the predictions to either match the performance of an ideal expected utility investor or a regret minimizing investor.The BoB-OPS algorithm has $O(\log T)$ regret with respect to the expected utility investor and $O(n \log T)$ static regret.

The OEU-LB-FTRL algorithm with log utility and previous return as the prediction gives a gradual variation bound of $O(\sqrt{n\tilde{V}'_T \log T})$. Finding an algorithm with logarithmic gradual variation measured in V_T and bounded worst case regret remains open problem.

References

[1] T. M. Cover, "Universal portfolios," *Mathematical Finance*, vol. 1, no. 1, pp. 1–29, 1991. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9965.1991.tb00002.x.

[2] T. Van Erven, D. Van der Hoeven, W. Kotłowski, and W. M. Koolen, "Open problem: Fast and optimal online portfolio selection," in *Proceedings of Thirty Third Conference on Learning Theory*, J. Abernethy and S. Agarwal, Eds., ser. Proceedings of Machine Learning Research, vol. 125, PMLR, 2020, pp. 3864–3869.

[3] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Mach. Learn.*, vol. 69, no. 2-3, pp. 169–192, 2007.

[4] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006, ISBN: 978-0-521-84108-5.

[5] N. Cesa-Bianchi and F. Orabona, "Online learning algorithms," *Annual Review of Statistics and Its Application*, vol. 8, no. 1, pp. 165–190, 2021.

[6] E. Hazan, "Introduction to online convex optimization," *Found. Trends Optim.*, vol. 2, no. 3-4, pp. 157–325, 2016.

[7] S. C. Hoi, D. Sahoo, J. Lu, and P. Zhao, "Online learning: A comprehensive survey," *Neurocomputing*, vol. 459, pp. 249–289, 2021.

[8] F. Orabona, "A modern introduction to online learning," *CoRR*, vol. abs/1912.13213, 2019. arXiv: 1912.13213.

[9] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, 2012.

[10] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, T. Fawcett and N. Mishra, Eds., AAAI Press, 2003, pp. 928–936.

[11] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM J. Comput.*, vol. 32, no. 1, pp. 48–77, 2002.

[12] A. Rakhlin and K. Sridharan, "Online learning with predictable sequences," in *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University,*

NJ, USA, S. Shalev-Shwartz and I. Steinwart, Eds., ser. JMLR Workshop and Conference Proceedings, vol. 30, JMLR.org, 2013, pp. 993–1019.

[13] F. Orabona and D. Pál, "Scale-free online learning," *Theor. Comput. Sci.*, vol. 716, pp. 50–69, 2018.

[14] J. C. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, 2011.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[16] R. Huang, T. Lattimore, A. György, and C. Szepesvári, "Following the leader and fast rates in linear prediction: Curved constraint sets and other regularities," in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 4970–4978.

[17] R. Huang, T. Lattimore, A. György, and C. Szepesvári, "Following the leader and fast rates in online linear prediction: Curved constraint sets and other regularities," *Journal of Machine Learning Research*, vol. 18, no. 145, pp. 1–31, 2017.

[18] J. Kivinen and M. K. Warmuth, "Exponentiated gradient versus gradient descent for linear predictors," *Inf. Comput.*, vol. 132, no. 1, pp. 1–63, 1997.

[19] A. Rakhlin and K. Sridharan, "Optimization, learning, and games with predictable sequences," in *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, Eds., 2013, pp. 3066–3074.

[20] K. S. Azoury and M. K. Warmuth, "Relative loss bounds for on-line density estimation with the exponential family of distributions," *Mach. Learn.*, vol. 43, no. 3, pp. 211–246, 2001.

[21] C. Chiang *et al.*, "Online optimization with gradual variations," in *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland*, S. Mannor, N. Srebro, and R. C. Williamson, Eds., ser. JMLR Proceedings, vol. 23, JMLR.org, 2012, pp. 6.1–6.20.

[22] J. Steinhardt and P. Liang, "Adaptivity and optimism: An improved exponentiated gradient algorithm," in *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, ser. JMLR Workshop and Conference Proceedings, vol. 32, JMLR.org, 2014, pp. 1593–1601.

[23] C. Wei and H. Luo, "More adaptive algorithms for adversarial bandits," in *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, S. Bubeck, V. Perchet, and P. Rigollet, Eds., ser. Proceedings of Machine Learning Research, vol. 75, PMLR, 2018, pp. 1263–1291.

[24] V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire, "Fast convergence of regularized learning in games," in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds., 2015, pp. 2989–2997.

[25] M. Mitzenmacher and S. Vassilvitskii, "Algorithms with predictions," *Commun. ACM*, vol. 65, no. 7, pp. 33–35, 2022.

[26] H. Luo, *Csci 659 lecture 3 : Adaptive algorithms and optimistic ftrl*, 2022.

[27] G. Farina, I. Anagnostides, H. Luo, C. Lee, C. Kroer, and T. Sandholm, "Near-optimal no-regret learning dynamics for general convex games," in *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., 2022.

[28] W. M. Koolen, *Adaftrl*, 2016.

[29] H. Hadiji and G. Stoltz, "Adaptation to the range in k-armed bandits," *J. Mach. Learn. Res.*, vol. 24, 13:1–13:33, 2023.

[30] Y. E. Nesterov, *Lectures on convex optimization* (Applied Optimization). Springer, 2018, vol. 137.

[31] J. Audibert and S. Bubeck, "Minimax policies for adversarial and stochastic bandits," in *COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009*, 2009.

[32] J. Audibert, S. Bubeck, and G. Lugosi, "Minimax policies for combinatorial prediction games," in *COLT 2011 - The 24th Annual Conference on Learning Theory, June 9-11, 2011, Budapest, Hungary*, S. M. Kakade and U. von Luxburg, Eds., ser. JMLR Proceedings, vol. 19, JMLR.org, 2011, pp. 107–132.

[33] J. Audibert, S. Bubeck, and G. Lugosi, "Regret in online combinatorial optimization," *Math. Oper. Res.*, vol. 39, no. 1, pp. 31–45, 2014.

[34] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 1–122, 2012.

[35] T. Lattimore and C. Szepesvári, *Bandit Algorithms*. Cambridge University Press, 2020.

[36] A. Slivkins, "Introduction to multi-armed bandits," *Found. Trends Mach. Learn.*, vol. 12, no. 1-2, pp. 1–286, 2019.

[37] S. de Rooij, T. van Erven, P. D. Grünwald, and W. M. Koolen, "Follow the leader if you can, hedge if you must," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1281–1316, 2014.

[38] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

[39] C. Allenberg, P. Auer, L. Györfi, and G. Ottucsák, "Hannan consistency in on-line learning in case of unbounded losses under partial monitoring," in *Algorithmic Learning Theory, 17th International Conference, ALT 2006, Barcelona, Spain, October 7-10, 2006, Proceedings*, J. L. Balcázar, P. M. Long, and F. Stephan, Eds., ser. Lecture Notes in Computer Science, vol. 4264, Springer, 2006, pp. 229–243.

[40] D. J. Foster, Z. Li, T. Lykouris, K. Sridharan, and É. Tardos, "Learning in games: Robustness of fast convergence," in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 4727–4735.

[41] R. Pogodin and T. Lattimore, "On first-order bounds, variance and gap-dependent bounds for adversarial bandits," in *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, A. Globerson and R. Silva, Eds., ser. Proceedings of Machine Learning Research, vol. 115, AUAI Press, 2019, pp. 894–904.

[42] S. Ito, "Parameter-free multi-armed bandit algorithms with hybrid data-dependent regret bounds," in *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, M. Belkin and S. Kpotufe, Eds., ser. Proceedings of Machine Learning Research, vol. 134, PMLR, 2021, pp. 2552–2583.

[43] E. Hazan and S. Kale, "Better algorithms for benign bandits," *J. Mach. Learn. Res.*, vol. 12, pp. 1287–1311, 2011.

[44] S. Bubeck, M. B. Cohen, and Y. Li, "Sparsity, variance and curvature in multi-armed bandits," in *Algorithmic Learning Theory, ALT 2018, 7-9 April 2018, Lanzarote, Canary Islands, Spain*, F. Janoos, M. Mohri, and K. Sridharan, Eds., ser. Proceedings of Machine Learning Research, vol. 83, PMLR, 2018, pp. 111–127.

[45] S. Bubeck, Y. Li, H. Luo, and C. Wei, "Improved path-length regret bounds for bandits," in *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, A.

Beygelzimer and D. Hsu, Eds., ser. Proceedings of Machine Learning Research, vol. 99, PMLR, 2019, pp. 508–528.

[46] J. Zimmert and Y. Seldin, "Tsallis-inf: An optimal algorithm for stochastic and adversarial bandits," *J. Mach. Learn. Res.*, vol. 22, 28:1–28:49, 2021.

[47] S. Gerchinovitz and T. Lattimore, "Refined lower bounds for adversarial bandits," in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 1190–1198.

[48] N. Cesa-Bianchi and O. Shamir, "Bandit regret scaling with the effective loss range," in *Algorithmic Learning Theory, ALT 2018, 7-9 April 2018, Lanzarote, Canary Islands, Spain*, F. Janoos, M. Mohri, and K. Sridharan, Eds., ser. Proceedings of Machine Learning Research, vol. 83, PMLR, 2018, pp. 128–151.

[49] T. Lattimore and C. Szepesvári, "Exploration by optimisation in partial monitoring," *CoRR*, vol. abs/1907.05772, 2019. arXiv: 1907.05772.

[50] M. Chen and X. Zhang, "Improved algorithms for adversarial bandits with unbounded losses," *arXiv preprint arXiv:2310.01756*, 2023.

[51] S. R. Putta and S. Agrawal, "Scale-free adversarial multi armed bandits," in *International Conference on Algorithmic Learning Theory, 29 March - 1 April 2022, Paris, France*, S. Dasgupta and N. Haghtalab, Eds., ser. Proceedings of Machine Learning Research, vol. 167, PMLR, 2022, pp. 910–930.

[52] T. M. Cover and E. Ordentlich, "Universal portfolios with side information," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 348–363, 1996.

[53] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth, "On-line portfolio selection using multiplicative updates," *Mathematical Finance*, vol. 8, no. 4, pp. 325–347, 1998. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9965.00058.

[54] C. Tsai, H. Cheng, and Y. Li, "Online self-concordant and relatively smooth minimization, with applications to online portfolio selection and learning quantum states," in *International Conference on Algorithmic Learning Theory, February 20-23, 2023, Singapore*, S. Agrawal and F. Orabona, Eds., ser. Proceedings of Machine Learning Research, vol. 201, PMLR, 2023, pp. 1481–1483.

[55] A. Kalai and S. S. Vempala, "Efficient algorithms for universal portfolios," *J. Mach. Learn. Res.*, vol. 3, pp. 423–440, 2002.

[56] L. Orseau, T. Lattimore, and S. Legg, "Soft-bayes: Prod for mixtures of experts with log-loss," in *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*, S. Hanneke and L. Reyzin, Eds., ser. Proceedings of Machine Learning Research, vol. 76, PMLR, 2017, pp. 372–399.

[57] H. Luo, C. Wei, and K. Zheng, "Efficient online portfolio with logarithmic regret," in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., 2018, pp. 8245–8255.

[58] J. Zimmert, N. Agarwal, and S. Kale, "Pushing the efficiency-regret pareto frontier for online learning of portfolios and quantum states," in *Conference on Learning Theory, 2-5 July 2022, London, UK*, P. Loh and M. Raginsky, Eds., ser. Proceedings of Machine Learning Research, vol. 178, PMLR, 2022, pp. 182–226.

[59] Z. Mhammedi and A. Rakhlin, "Damped online newton step for portfolio selection," in *Conference on Learning Theory, 2-5 July 2022, London, UK*, P. Loh and M. Raginsky, Eds., ser. Proceedings of Machine Learning Research, vol. 178, PMLR, 2022, pp. 5561–5595.

[60] R. Jézéquel, D. M. Ostrovskii, and P. Gaillard, "Efficient and near-optimal online portfolio selection," *arXiv preprint arXiv:2209.13932*, 2022.

[61] A. Agarwal and E. Hazan, "Efficient algorithms for online game playing and universal portfolio management," *Electron. Colloquium Comput. Complex.*, vol. TR06-033, 2006. ECCC: TR06-033.

[62] E. Hazan and S. Kale, "An online portfolio selection algorithm with regret logarithmic in price variation," *Mathematical Finance*, vol. 25, no. 2, pp. 288–310, 2015. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mafi.12006.

[63] T. van Erven, P. Grunwald, W. M. Koolen, and S. de Rooij, "Adaptive hedge," in *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, Eds., 2011, pp. 1656–1664.

[64] T. van Erven, W. M. Koolen, and D. van der Hoeven, "Metagrad: Adaptation using multiple learning rates in online learning," *J. Mach. Learn. Res.*, vol. 22, 161:1–161:61, 2021.

[65] G. Wang, S. Lu, and L. Zhang, "Adaptivity and optimality: A universal algorithm for online convex optimization," in *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, A. Globerson and R. Silva, Eds., ser. Proceedings of Machine Learning Research, vol. 115, AUAI Press, 2019, pp. 659–668.

[66] F. Orabona, N. Cesa-Bianchi, and C. Gentile, "Beyond logarithmic bounds in online learning," in *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012, La Palma, Canary Islands, Spain, April 21-23, 2012*, N. D. Lawrence and M. A. Girolami, Eds., ser. JMLR Proceedings, vol. 22, JMLR.org, 2012, pp. 823–831.

[67] Y. Yan, P. Zhao, and Z. Zhou, "Universal online learning with gradient variations: A multi-layer online ensemble approach," in *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., 2023.

[68] C. Tsai, Y. Lin, and Y. Li, "Data-dependent bounds for online portfolio selection without lipschitzness and smoothness," in *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., 2023.

[69] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[70] J. L. Kelly, "A new interpretation of information rate," *the bell system technical journal*, vol. 35, no. 4, pp. 917–926, 1956.

[71] E. O. Thorp, "Portfolio choice and the kelly criterion," in *Stochastic Optimization Models in Finance*, W. ZIEMBA and R. VICKSON, Eds., Academic Press, 1975, pp. 599–619, ISBN: 978-0-12-780850-5.

[72] N. Srebro, K. Sridharan, and A. Tewari, "Smoothness, low noise and fast rates," in *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada*, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Eds., Curran Associates, Inc., 2010, pp. 2199–2207.

Appendix A: Proofs from Chapter 2

We begin by stating a few useful properties of Bregman Divergences.

Lemma A.1. *For any $v, w \in dom(\nabla F)$ and $u \in dom(F)$ we have:*

$$B_F(u\|w) - B_F(u\|v) - B_F(v\|w) = (\nabla F(w) - \nabla F(v))^\top (v - u)$$

Lemma A.1 is called the *Law of cosines for Bregman divergence*. The proof is via a direct calculation. The law of cosines can be extended to the case of Mixed Bregmans as well.

Lemma 1.2. *For any $v, w \in dom(\nabla F)$ and $u \in dom(F)$ we have:*

$$B_F^{a,b}(u\|w) - B_F^{a,c}(u\|v) - B_F^{c,b}(v\|w) = \left(\frac{\nabla F(w)}{b} - \frac{\nabla F(v)}{c}\right)^\top (v - u)$$

We now prove the main theorem obtaining a general regret bound for Optimistic FTRL.

Theorem 2.3. *For any $w \in \mathcal{D}$, any sequence of convex cost functions $f_1, \ldots, f_T$, convex hint functions $m_1, \ldots, m_T$, convex regularizer F and parameters $\eta_0, \ldots, \eta_T$ such that $w_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + m_t(w) + \frac{F(w)}{\eta_{t-1}}$ and $w'_t \in \arg\min_{w\in\mathcal{D}} \sum_{s=1}^{t-1} f_s(w) + \frac{F(w)}{\eta_{t-1}}$. Let $g_t = \sum_{s=1}^{t} f_s$. The iterates of Optimistic FRTL $w_1, \ldots, w_T$ satisfies the regret inequality $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:*

$$\leq B_F^{\eta_T,\eta_0}(w\|w'_1) + \sum_{t=1}^{T} \Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - B_{g_t}(w'_{t+1}\|w_t) - B_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t) \\ - B_{g_{t-1}}(w_t\|w'_t) - B_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t) \Big]$$

Further, if F is such that $\min_{w\in\mathcal{D}} F(w) = 0$ and the sequence $\eta_0, \ldots, \eta_T$ is non-increasing, then

the above bound simplifies to $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$*:*

$$\leq \frac{F(w)}{\eta_T} + \sum_{t=1}^{T} \Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - B_{g_t}(w'_{t+1} \| w_t) - \frac{B_F(w'_{t+1} \| w_t)}{\eta_{t-1}} \\ - B_{g_{t-1}}(w_t \| w'_t) - \frac{B_F(w_t \| w'_t)}{\eta_{t-1}} \Big]$$

Proof. Consider $f_t(w_t) - f_t(w)$. We expand it by adding and subtracting $f_t(w'_{t+1})$.

$$f_t(w_t) - f_t(w) = f_t(w_t) - f_t(w'_{t+1}) + f_t(w'_{t+1}) - f_t(w)$$

Using the definition of Bregman Divergence B_{f_t}

$$= \nabla f_t(w_t)^\top (w_t - w'_{t+1}) - \mathrm{B}_{f_t}(w'_{t+1} \| w_t) + \underbrace{\nabla f_t(w'_{t+1})^\top (w'_{t+1} - w)}_{(1)} - \mathrm{B}_{f_t}(w \| w'_{t+1})$$

Consider term (1) in the above equation. Let $g_t(w) = \sum_{s=1}^{t} f_s(w)$. We can write $f_t(w) = g_t(w) - g_{t-1}(w)$, so $\nabla f_t(w) = \nabla g_t(w) - \nabla f_{t-1}(w)$. Substituting this expression for $\nabla f_t(w'_{t+1})$

$$(1) = \nabla f_t(w'_{t+1})^\top (w'_{t+1} - w) = (\nabla g_t(w'_{t+1}) - \nabla g_{t-1}(w'_{t+1}))^\top (w'_{t+1} - w)$$

Adding and subtracting $\nabla g_{t-1}(w_t)^\top (w'_{t+1} - w)$

$$= (\nabla g_t(w'_{t+1}) - \nabla g_{t-1}(w_t))^\top (w'_{t+1} - w) + \underbrace{(\nabla g_{t-1}(w_t) - \nabla g_{t-1}(w'_{t+1}))^\top (w'_{t+1} - w)}_{(2)}$$

Using Lemma A.1, the term (2) is:

$$(\nabla g_{t-1}(w_t) - \nabla g_{t-1}(w'_{t+1}))^\top (w'_{t+1} - w) = \mathrm{B}_{g_{t-1}}(w \| w_t) - \mathrm{B}_{g_{t-1}}(w \| w'_{t+1}) - \mathrm{B}_{g_{t-1}}(w'_{t+1} \| w_t)$$

Substituting this back in the expression for $f_t(w_t) - f_t(w)$ and rearranging, we have:

$$\begin{aligned}
f_t(w_t) - f_t(w) &= \nabla f_t(w_t)^\top (w_t - w'_{t+1}) - \mathrm{B}_{f_t}(w'_{t+1} \| w_t) \\
&\quad + (\nabla g_t(w'_{t+1}) - \nabla g_{t-1}(w_t))^\top (w'_{t+1} - w) \\
&\quad + \mathrm{B}_{g_{t-1}}(w \| w_t) - \mathrm{B}_{g_{t-1}}(w \| w'_{t+1}) - \mathrm{B}_{g_{t-1}}(w'_{t+1} \| w_t) \\
&\quad - \mathrm{B}_{f_t}(w \| w'_{t+1}) \\
&= \nabla f_t(w_t)^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1} \| w_t) \\
&\quad + (\nabla g_t(w'_{t+1}) - \nabla g_{t-1}(w_t))^\top (w'_{t+1} - w) \\
&\quad + \mathrm{B}_{g_{t-1}}(w \| w_t) - \mathrm{B}_{g_t}(w \| w'_{t+1}) \\
&= (\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1} \| w_t) \\
&\quad + (\nabla g_t(w'_{t+1}) - \nabla g_{t-1}(w_t) - \nabla m_t(w_t))^\top (w'_{t+1} - w) \\
&\quad + \mathrm{B}_{g_{t-1}}(w \| w_t) - \mathrm{B}_{g_t}(w \| w'_{t+1}) - \nabla m_t(w_t)^\top (w - w_t)
\end{aligned}$$

By Lemma A.1, we can write:

$$\mathrm{B}_{g_{t-1}}(w \| w_t) = \mathrm{B}_{g_{t-1}}(w \| w'_t) - \mathrm{B}_{g_{t-1}}(w_t \| w'_t) + (\nabla g_{t-1}(w'_t) - \nabla g_{t-1}(w_t))^\top (w - w_t)$$

Substituting this back in the expression for $f_t(w_t) - f_t(w)$, we and simplifying, we have:

$$\begin{aligned}
f_t(w_t) - f_t(w) &= (\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1} \| w_t) - \mathrm{B}_{g_{t-1}}(w_t \| w'_t) \\
&\quad + \underbrace{(\nabla g_t(w'_{t+1}) - \nabla g_{t-1}(w_t) - \nabla m_t(w_t))^\top (w'_{t+1} - w)}_{(3)} \\
&\quad + \underbrace{(\nabla g_{t-1}(w_t) + \nabla m_t(w_t) - \nabla g_{t-1}(w'_t))^\top (w_t - w)}_{(4)} \\
&\quad + \mathrm{B}_{g_{t-1}}(w \| w'_t) - \mathrm{B}_{g_t}(w \| w'_{t+1})
\end{aligned}$$

We introduce the following notation to ease the algebraic manipulation:

$$G_{t-1}(w) = g_{t-1}(w) + m_t(w) + \frac{F(w)}{\eta_{t-1}}$$
$$G'_t(w) = g_t(w) + \frac{F(w)}{\eta_t}$$

We simplify term (3) and apply Lemma 1.2:

$$\begin{aligned}(3) &= (\nabla G'_t(w'_{t+1}) - \nabla G_{t-1}(w_t))^\top (w'_{t+1} - w) + \left(\frac{\nabla F(w_t)}{\eta_{t-1}} - \frac{\nabla F(w'_{t+1})}{\eta_t}\right)^\top (w'_{t+1} - w) \\ &= (\nabla G'_t(w'_{t+1}) - \nabla G_{t-1}(w_t))^\top (w'_{t+1} - w) + \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w_t) - \mathrm{B}_F^{\alpha,\eta_t}(w\|w'_{t+1}) - \mathrm{B}_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t)\end{aligned}$$

Similarly, we simplify term (4) and apply Lemma 1.2:

$$\begin{aligned}(4) &= (\nabla G_{t-1}(w_t) - \nabla G'_{t-1}(w'_t))^\top (w_t - w) + \left(\frac{\nabla F(w'_t)}{\eta_{t-1}} - \frac{\nabla F(w_t)}{\eta_{t-1}}\right)^\top (w_t - w) \\ &= (\nabla G_{t-1}(w_t) - \nabla G'_{t-1}(w'_t))^\top (w_t - w) + \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w'_t) - \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w_t) - \mathrm{B}_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t)\end{aligned}$$

Substituting these back in the expression for $f_t(w_t) - f_t(w)$, we have:

$$\begin{aligned}&= (\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1}\|w_t) - \mathrm{B}_{g_{t-1}}(w_t\|w'_t) \\ &\quad + (\nabla G'_t(w'_{t+1}) - \nabla G_{t-1}(w_t))^\top (w'_{t+1} - w) + \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w_t) - \mathrm{B}_F^{\alpha,\eta_t}(w\|w'_{t+1}) - \mathrm{B}_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t) \\ &\quad + (\nabla G_{t-1}(w_t) - \nabla G'_{t-1}(w'_t))^\top (w_t - w) + \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w'_t) - \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w_t) - \mathrm{B}_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t) \\ &\quad + \mathrm{B}_{g_{t-1}}(w\|w'_t) - \mathrm{B}_{g_t}(w\|w'_{t+1}) \\ &= (\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1}\|w_t) - \mathrm{B}_{g_{t-1}}(w_t\|w'_t) \\ &\quad - \mathrm{B}_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t) - \mathrm{B}_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t) \\ &\quad + \mathrm{B}_{g_{t-1}}(w\|w'_t) - \mathrm{B}_{g_t}(w\|w'_{t+1}) + \mathrm{B}_F^{\alpha,\eta_{t-1}}(w\|w'_t) - \mathrm{B}_F^{\alpha,\eta_t}(w\|w'_{t+1}) \\ &\quad + (\nabla G'_{t-1}(w'_t) - \nabla G'_t(w'_{t+1}))^\top w \\ &\quad + \nabla G_{t-1}(w_t)^\top (w_t - w'_{t+1})\end{aligned}$$

$$+ \nabla G'_t(w'_{t+1})w'_{t+1} - \nabla G'_{t-1}(w'_t)^\top w_t$$

Since w_t minimizes $G_{t-1}(w)$, we have $(\nabla G_{t-1}(w_t))^\top (w_t - w'_{t+1}) \leq 0$. Taking the summation over the t terms $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$, we have :

$$\begin{aligned}
&\leq \sum_{t=1}^{T} \left((\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1} \| w_t) - \mathrm{B}_F^{\eta_t, \eta_{t-1}}(w'_{t+1} \| w_t) \right) \\
&\quad + \sum_{t=1}^{T} \left(-\mathrm{B}_{g_{t-1}}(w_t \| w'_t) - \mathrm{B}_F^{\eta_{t-1}, \eta_{t-1}}(w_t \| w'_t) \right) \\
&\quad + \underbrace{\sum_{t=1}^{T} \mathrm{B}_{g_{t-1}}(w \| w'_t) - \mathrm{B}_{g_t}(w \| w'_{t+1})}_{(5)} + \underbrace{\sum_{t=1}^{T} \mathrm{B}_F^{\alpha, \eta_{t-1}}(w \| w'_t) - \mathrm{B}_F^{\alpha, \eta_t}(w \| w'_{t+1})}_{(6)} \\
&\quad + \underbrace{\sum_{t=1}^{T} (\nabla G'_{t-1}(w'_t) - \nabla G'_t(w'_{t+1}))^\top w}_{(7)} + \underbrace{\sum_{t=1}^{T} \left(\nabla G'_t(w'_{t+1})w'_{t+1} - \nabla G'_{t-1}(w'_t)^\top w_t \right)}_{(8)}
\end{aligned}$$

We can telescope term (5) to get:

$$\sum_{t=1}^{T} \mathrm{B}_{g_{t-1}}(w \| w'_t) - \mathrm{B}_{g_t}(w \| w'_{t+1}) = \mathrm{B}_{g_0}(w \| w'_1) - \mathrm{B}_{g_T}(w \| w'_{T+1}) = 0 - \mathrm{B}_{g_T}(w \| w'_{T+1}) \leq 0$$

We can telescope term (6) to get:

$$\sum_{t=1}^{T} \mathrm{B}_F^{\alpha, \eta_{t-1}}(w \| w'_t) - \mathrm{B}_F^{\alpha, \eta_t}(w \| w'_{t+1}) = \mathrm{B}_F^{\alpha, \eta_0}(w \| w'_1) - \mathrm{B}_F^{\alpha, \eta_T}(w \| w'_{T+1})$$

Taking $\alpha = \eta_T$, we have:

$$\mathrm{B}_F^{\eta_T, \eta_0}(w \| w'_1) - \mathrm{B}_F^{\eta_T, \eta_T}(w \| w'_{T+1}) \leq \mathrm{B}_F^{\eta_T, \eta_0}(w \| w'_1)$$

Term (7) can be telescoped as:

$$\sum_{t=1}^{T}(\nabla G'_{t-1}(w'_t) - \nabla G'_t(w'_{t+1}))^\top w = (\nabla G'_0(w'_1) - \nabla G'_T(w'_{T+1}))^\top w = -\nabla G_T(w'_{T+1})^\top w$$

The hint for round $T+1$ can be taken as $m_{T+1}(w) = 0$. We have $w_{T+1} = w'_{T+1}$. Finally for term (8):

$$\begin{aligned}\sum_{t=1}^{T} \nabla G'_t(w'_{t+1})^\top w'_{t+1} - \nabla G'_{t-1}(w'_t)^\top w_t &= \sum_{t=1}^{T-1} \nabla G'_t(w'_{t+1})^\top (w'_{t+1} - w_{t+1}) + \nabla G'_T(w'_{T+1})^\top w_{T+1} \\ &\leq \nabla G'_T(w'_{T+1})^\top w'_{T+1}\end{aligned}$$

Here, we used the fact that w'_{t+1} minimizes $G'_t(w)$. So $\nabla G'_t(w'_{t+1})^\top (w'_{t+1} - w) \leq 0$ for all $w \in \mathcal{D}$. Combining the upper bounds for terms (7) and (8):

$$(7) + (8) \leq \nabla G_T(w'_{T+1})^\top (w'_{T+1} - w) \leq 0$$

Thus, we have the result $\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq$:

$$\begin{aligned}\leq \mathrm{B}_F^{\eta_T,\eta_0}(w\|w'_1) + \sum_{t=1}^{T}\Big[&(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1}\|w_t) - \mathrm{B}_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t) \\ &- \mathrm{B}_{g_{t-1}}(w_t\|w'_t) - \mathrm{B}_F^{\eta_{t-1},\eta_{t-1}}(w_t\|w'_t)\Big]\end{aligned}$$

Further, if F is such that $min_{w\in\mathcal{D}} F(w) = 0$, then $w'_1 \in \min_{w\in\mathcal{D}} F(w)$. So, $\mathrm{B}_F^{\eta_T,\eta_0}(w\|w'_1) \leq \frac{F(w)}{\eta_T}$. If the sequence $\eta_0, \ldots, \eta_T$ is non-increasing, then

$$\mathrm{B}_F^{\eta_t,\eta_{t-1}}(w'_{t+1}\|w_t) \geq \frac{1}{\eta_{t-1}}\mathrm{B}_F(w'_{t+1}\|w_t)$$

the above bound simplifies to $\sum_{t=1}^{T} f_t(w_t) - f_t(w)$:

$$\leq \frac{F(w)}{\eta_T} + \sum_{t=1}^{T}\Big[(\nabla f_t(w_t) - \nabla m_t(w_t))^\top (w_t - w'_{t+1}) - \mathrm{B}_{g_t}(w'_{t+1}\|w_t) - \frac{\mathrm{B}_F(w'_{t+1}\|w_t)}{\eta_{t-1}}$$

$$- \mathrm{B}_{g_{t-1}}(w_t \| w'_t) - \frac{\mathrm{B}_F(w_t \| w'_t)}{\eta_{t-1}} \Big]$$

■

Appendix B: Proofs from Chapter 4

Let I be the $n \times n$ identity matrix. We state the following lemma, which is a tighter version of Lemma 11 in Hazan *et al.* [3]

Lemma B.1. *[3, Lemma 11] Let $x_1, \ldots, x_t$ be a sequence of vectors in $\mathbb{R}^n$. Define $H_t = \epsilon I + \sum_{s=1}^{t} x_s x_s^\top$. Then, the following holds:*

$$\sum_{t=1}^{T} x_t^\top H_t^{-1} x_t \leq n \log \left(1 + \frac{\sum_{t=1}^{T} \|x_t\|_2^2}{n\epsilon}\right)$$

Lemma B.2. *For any $w_1, \ldots, w_T \in \Delta_n$, and $r_1, \ldots, r_T \in \mathbb{R}_+^n$ we have the inequality:*

$$\frac{1}{2} \sum_{t=1}^{T} \frac{r_t}{r_t^\top w_t}^\top \left(\sum_{s=1}^{t} \frac{r_s r_s^\top}{(r_s^\top w_s)(\max_i r_s(i))} + \epsilon I + \lambda I\right)^{-1} \frac{r_t}{r_t^\top w_t} \leq \frac{nR}{2} \log \left(1 + \frac{\sum_{t=1}^{T} \|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)}\right)$$

Here $R = \max_{t,i,j} \frac{r_t(i)}{r_t(j)}$ and $\hat{r}_s = \frac{r_s}{\sqrt{(r_s^\top w_s)(\max_i r_s(i))}}$

Proof. Let $\tilde{r}_t = \frac{r_t}{\max_i r_t(i)}$. We re-write the above expression with $\tilde{r}_t$ as:

$$\frac{1}{2} \sum_{t=1}^{T} \frac{1}{(\tilde{r}_t^\top w_t)^2} \tilde{r}_t^\top \left(\sum_{s=1}^{t} \frac{\tilde{r}_s \tilde{r}_s^\top}{\tilde{r}_s^\top w_s} + \epsilon I + \lambda I\right)^{-1} \tilde{r}_t$$

Now let $\hat{r}_t = \frac{\tilde{r}_t}{\sqrt{\tilde{r}_t^\top w_t}} = \frac{r_t}{\sqrt{(r_t^\top w_t)(\max_i r_t(i))}} = \frac{r_t}{r_t^\top w_t} \sqrt{\frac{r_t^\top w_t}{\max_i r_t(i)}}$

$$\frac{1}{2 \min_t (\tilde{r}_t^\top w_t)} \sum_{t=1}^{T} \hat{r}_t \left(\sum_{s=1}^{t} \hat{r}_s \hat{r}_s^\top + \epsilon I + \lambda I\right)^{-1} \hat{r}_t$$

We have $\frac{1}{\min_t (\tilde{r}_t^\top w_t)} = \max_t \frac{1}{(\tilde{r}_t^\top w_t)} = \max_t \frac{\max_i r_t(i)}{r_t^\top w_t} \leq \max_t \frac{\max_i r_t(i)}{\min_i r_t(i)} = \max_{t,i,j} \frac{r_t(i)}{r_t(j)} = R$. Using the

so called Elliptical potential lemma (Lemma B.1), we have the bound:

$$\sum_{t=1}^{T} \hat{r}_t \left(\sum_{s=1}^{t} \hat{r}_s \hat{r}_s^\top + \epsilon I + \lambda I \right)^{-1} \hat{r}_t \leq n \log \left(1 + \frac{\sum_{t=1}^{T} \|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)} \right)$$

This gives us the stated result. ■

Lemma B.3. *[72, Lemma 3.1] If a non-negative function f is H-smooth on the domain $\mathcal{D}$, then $\|\nabla f(w)\| \leq \sqrt{4Hf(w)}$ for all $w \in \mathcal{D}$*

Lemma B.4. *Let $f_t = -\log(r_t^\top w)$ and let $w_t^\star = \arg\min_{w \in \Delta_n} f_t(w)$, i.e., it is the optimal portfolio for the return vector r_t. Let $l_t(w) = f_t(w) - f_t(w_t^\star)$. Then l_t is nR^2-smooth on Δ_n. So,*

$$\|\nabla f_t(w)\|_2^2 \leq 4nR^2 l_t(w)$$

Proof. We have $\nabla l_t(w) = \nabla f_t(w)$. So for any $w, w' \in \Delta_n$, we have:

$$s\|\nabla l_t(w) - \nabla l_t(w')\|_2 = \left\| \frac{r_t}{r_t^\top w} - \frac{r_t}{r_t^\top w'} \right\|_2 = \frac{\|r_t\|_2^2 \|w - w'\|_2}{(r_t^\top w)(r_t^\top w')} \leq nR^2 \|w - w'\|_2$$

Thus $l_t(w)$ is nR^2 smooth on Δ_n. Applying Lemma B.3, we have the final result. ■

Lemma B.5. *[68, Lemma 4.7] Let $f_t = -\log(r_t^\top w)$ and let $w_t^\star = \arg\min_{w \in \Delta_n} f_t(w)$, i.e., it is the optimal portfolio for the return vector r_t. Let $l_t(w) = f_t(w) - f_t(w_t^\star)$. We have,*

$$\inf_c \|(\nabla f_t(w) + c\mathbf{1}) \circ w\|_2^2 \leq 4 l_t(w)$$

Lemma B.6. *[66, Corollary 5] Let $a, b, c, d, x > 0$ satisfy $x \leq a \log(bx + c) + d$, then:*

$$x \leq a \log \left(2 \left(ab \log \left(\frac{2ab}{e} \right) + db + c \right) \right) + d$$

Here e is the base of the natural logarithm.

Lemma B.7. *[8, Lemma 4.24] Let $a, b, c, , x > 0$ satisfy $x \leq c + \sqrt{ax + b}$, then:*

$$x \leq a + c + 2\sqrt{b + ac}$$

Theorem 4.4. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}^n_+$, define $f_t(w) = -\log(r_t^\top w)$. The updates of AdaCurv ONS (Equation (4.5)) with $\epsilon = 1$ satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{1}{2} + \frac{nR}{2} \log\left(4nR^3 \log\left(\frac{4nR^3}{e}\right) + 4R^2 + 8R^2 L_T^\star + 2\right)$$

Here, $L_T^\star = \min_{w \in \Delta_n} \left[\sum_{t=1}^T f_t(w)\right] - \sum_{t=1}^T \left[\min_{w \in \Delta_n} f_t(w)\right]$ is the regret between the best static and the best dynamic portfolio selection strategies.

Proof. Consider the result from Theorem 4.2:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{\epsilon}{2} + \inf_{\lambda \geq 0}\left(\lambda T + \frac{nR}{2} \log\left(1 + \frac{\sum_{t=1}^T \|\hat{r}_t\|_2^2}{n(\epsilon + \lambda)}\right)\right)$$

Note that :

$$\|\hat{r}_t\|_2^2 = \left\| \frac{r_t}{r_t^\top w_t} \sqrt{\frac{r_t^\top w_t}{\max_i r_t(i)}} \right\|_2^2 = \frac{r_t^\top w_t}{\max_i r_t(i)} \|\nabla f_t(w_t)\|_2^2 \leq \|\nabla f_t(w_t)\|_2^2$$

Let $w_t^\star = \arg\min_{w \in \Delta_n} f_t(w)$, i.e., it is the optimal portfolio for the return vector r_t. Let $l_t(w) = f_t(w) - f_t(w_t^\star)$. Note that $\nabla f_t(w) = \nabla l_t(w)$. So, we have:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{\epsilon}{2} + \inf_{\lambda \geq 0}\left(\lambda T + \frac{nR}{2} \log\left(1 + \frac{\sum_{t=1}^T \|\nabla f_t(w_t)\|_2^2}{n(\epsilon + \lambda)}\right)\right)$$

Pick $\epsilon = 1$ and $\lambda = 0$:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{1}{2} + \frac{nR}{2} \log\left(1 + \frac{\sum_{t=1}^T \|\nabla f_t(w_t)\|_2^2}{n}\right)$$

Apply Lemma B.4, which states that $\|\nabla f_t(w_t)\|_2^2 \leq 4nR^2 l_t(w_t) = 4nR^2 \left(f_t(w_t) - f_t(w_t^\star)\right)$:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{1}{2} + \frac{nR}{2} \log\left(1 + 4R^2\left(\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star)\right)\right)$$
$$\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star) \leq \sum_{t=1}^{T} f_t(w) - f_t(w_t^\star) + \frac{1}{2} + \frac{nR}{2} \log\left(1 + 4R^2\left(\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star)\right)\right)$$

Now, we apply Lemma B.6 with $a = nR/2$, $b = 4R^2$, $c = 1$, $d = 1/2 + \sum_{t=1}^{T} f_t(w) - f_t(w_t^\star)$ and $x = \sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star)$. So, we have the inequality:

$$\begin{aligned}\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star) \leq & \frac{1}{2} + \sum_{t=1}^{T} f_t(w) - f_t(w_t^\star) \\ & + \frac{nR}{2} \log\left(4nR^3 \log\left(\frac{4nR^3}{e}\right) + 4R^2 + 8R^2\left(\sum_{t=1}^{T} f_t(w) - f_t(w_t^\star)\right) + 2\right) \\ \sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq & \frac{1}{2} + \frac{nR}{2} \log\left(4nR^3 \log\left(\frac{4nR^3}{e}\right) + 4R^2 + 8R^2\left(\sum_{t=1}^{T} f_t(w) - f_t(w_t^\star)\right) + 2\right)\end{aligned}$$

Specifically, if we pick $w^\star \in \arg\min_{w \in \Delta_n} \sum_{t=1}^{T} f_t(w)$. Then, $\sum_{t=1}^{T} f_t(w^\star) - f_t(w_t^\star)$ is the regret between the best static and the best dynamic portfolio selection strategies. We use the shorthand $L_T^\star$ to denote this quantity. Thus, we have the bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{1}{2} + \frac{nR}{2} \log\left(4nR^3 \log\left(\frac{4nR^3}{e}\right) + 4R^2 + 8R^2 L_T^\star + 2\right)$$

■

Theorem 4.5. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, define $f_t(w) = -\log(r_t^\top w)$. The updates of LB-AdaCurv ONS (Algorithm 2) with $\epsilon = 1$ satisfy the regret bound:*

$$\begin{aligned}\sum_{t=1}^{T} f_t(w_t) - f_t(w_t) \leq & \frac{5}{2} + 2n \log T + \min\left[2 + 2\sqrt{8n \log T} + 4\sqrt{8n\left(L_T^\star + \frac{9}{2} + 2n \log T\right) \log T},\right. \\ & \left. nR \log\left(8nR^3 \log\left(\frac{8nR^3}{e}\right) + 20R^2 + 16R^2 n \log T + 8R^2 L_T^\star + 2\right)\right]\end{aligned}$$

Here, $L_T^\star = \min_{w\in\Delta_n}\left[\sum_{t=1}^T f_t(w)\right] - \sum_{t=1}^T\left[\min_{w\in\Delta_n} f_t(w)\right]$ *is the regret between the best static and the best dynamic portfolio selection strategies.*

Proof. From Theorem 4.3, we have the following bound for LB-AdaCurv ONS after picking $\epsilon = 1$ and $\lambda = 0$:

$$\sum_{t=1}^T f_t(w_t) - f_t(w) \le \frac{5}{2} + 2n\log T + 2\min\left[\frac{nR}{2}\log\left(1 + \frac{\sum_{t=1}^T \|\hat{r}_t\|_2^2}{n}\right), 1 + \sqrt{2n\left(\sum_{t=1}^T \inf_c \|(\nabla f_t(w_t) + c\mathbf{1})\circ w_t\|_2^2\right)\log(T)}\right]$$

Consider just the first part of the minimum. Note that :

$$\|\hat{r}_t\|_2^2 = \left\|\frac{r_t}{r_t^\top w_t}\sqrt{\frac{r_t^\top w_t}{\max_i r_t(i)}}\right\|_2^2 = \frac{r_t^\top w_t}{\max_i r_t(i)}\|\nabla f_t(w_t)\|_2^2 \le \|\nabla f_t(w_t)\|_2^2$$

Let $w_t^\star = \arg\min_{w\in\Delta_n} f_t(w)$, i.e., it is the optimal portfolio for the return vector r_t. Let $l_t(w) = f_t(w) - f_t(w_t^\star)$. Note that $\nabla f_t(w) = \nabla l_t(w)$. So, we have:

$$\sum_{t=1}^T f_t(w_t) - f_t(w) \le \frac{5}{2} + 2n\log T + nR\log\left(1 + \frac{\sum_{t=1}^T \|\nabla f_t(w_t\|_2^2}{n}\right)$$

Apply Lemma B.4, which states that $\|\nabla f_t(w_t)\|_2^2 \le 4nR^2 l_t(w_t) = 4nR^2\left(f_t(w_t) - f_t(w_t^\star)\right)$:

$$\sum_{t=1}^T f_t(w_t) - f_t(w) \le \frac{5}{2} + 2n\log T + nR\log\left(1 + 4R^2\left(\sum_{t=1}^T f_t(w_t) - f_t(w_t^\star)\right)\right.$$

$$\sum_{t=1}^T f_t(w_t) - f_t(w_t^\star) \le \sum_{t=1}^T f_t(w) - f_t(w_t^\star) + \frac{5}{2} + 2n\log T + nR\log\left(1 + 4R^2\left(\sum_{t=1}^T f_t(w_t) - f_t(w_t^\star)\right)\right)$$

Now, we apply Lemma B.6 with $a = nR$, $b = 4R^2$, $c = 1$, $d = 5/2 + 2n\log T + \sum_{t=1}^T f_t(w) - f_t(w_t^\star)$ and $x = \sum_{t=1}^T f_t(w_t) - f_t(w_t^\star)$. So, we have the inequality:

$$\sum_{t=1}^T f_t(w_t) - f_t(w_t^\star) \le \frac{5}{2} + 2n\log T + \sum_{t=1}^T f_t(w) - f_t(w_t^\star)$$

$$+ nR\log\left(8nR^3\log\left(\frac{8nR^3}{e}\right) + 20R^2 + 16R^2 n\log T + 8R^2\left(\sum_{t=1}^{T} f_t(w) - f_t(w_t^\star)\right) + 2\right)$$

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{5}{2} + 2n\log T +$$

$$+ nR\log\left(8nR^3\log\left(\frac{8nR^3}{e}\right) + 20R^2 + 16R^2 n\log T + 8R^2\left(\sum_{t=1}^{T} f_t(w) - f_t(w_t^\star)\right) + 2\right)$$

Specifically, if we pick $w^\star \in \arg\min_{w\in\Delta_n}\sum_{t=1}^{T} f_t(w)$. Then, $L_T^\star = \sum_{t=1}^{T} f_t(w^\star) - f_t(w_t^\star)$ is the regret between the best static and the best dynamic portfolio selection strategies. We use the shorthand $L_T^\star$ to denote this quantity. Thus, we have the bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{5}{2} + 2n\log T + nR\log\left(8nR^3\log\left(\frac{8nR^3}{e}\right) + 20R^2 + 16R^2 n\log T + 8R^2 L_T^\star + 2\right)$$

Consider the second part of the minimum:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{5}{2} + 2n\log T + 2 + 2\sqrt{2n\left(\sum_{t=1}^{T}\inf_c \|(\nabla f_t(w_t) + c\mathbf{1})\circ w_t\|_2^2\right)\log(T)}$$

Using Lemma B.5, we have the bound $\inf_c \|(\nabla f_t(w_t) + c\mathbf{1})\circ w_t\|_2^2 \leq 4(f_t(w_t) - f_t(w_t^\star))$:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{5}{2} + 2n\log T + 2 + 2\sqrt{8n\left(\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star)\right)\log(T)}$$

$$\implies \sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star) \leq \sum_{t=1}^{T} f_t(w) - f_t(w_t^\star) + \frac{5}{2} + 2n\log T + 2 + 2\sqrt{8n\left(\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star)\right)\log(T)}$$

Now, we apply Lemma B.7 with $c = \sum_{t=1}^{T} f_t(w) - f_t(w_t^\star) + \frac{5}{2} + 2n\log T + 2$, $a = 2\sqrt{8n\log T}$, $b = 0$:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w_t^\star) \leq \sum_{t=1}^{T} f_t(w) - f_t(w_t^\star) + \frac{5}{2} + 2n\log T + 2 + 2\sqrt{8n\log T} + 4\sqrt{8n\log T\left(\sum_{t=1}^{T} f_t(w) - f_t(w_t^\star) + \frac{5}{2} + 2n\log T + 2\right)}$$

$$\implies \sum_{t=1}^{T} f_t(w_t) - f_t(w) \leq \frac{9}{2} + 2n\log T + 2\sqrt{8n\log T} + 4\sqrt{8n\left(L_T^\star + \frac{9}{2} + 2n\log T\right)\log T}$$

Combining the two results, we get the final bound. ■

In order to obtain the $O(\log Q_T)$ regret bound, we state a slightly modified version of a theorem from Hazan and Kale [62].

Theorem B.8. *[62, Theorem 1.1] Let the cost functions be $f_t(w) = h_t(w^\top v_t)$ for a scalar function h_t. Consider the iterates:*

$$w_t = \arg\min_{w\in\mathcal{D}} \frac{1}{2}\|w\|_2^2 + \sum_{s=1}^{t-1} h_s(w^\top v_s)$$

If $\|v_t\| \leq V$, $\|w\| \leq D$ for all $w \in \mathcal{D}$, $h_t'(w_t^\top v_t) \in [-a, 0]$ and $h_t''(w^\top v_t) \geq b$ for all $w \in \mathcal{D}$, then:

$$\mathcal{R}_T(w) \leq O\left(\frac{a^2 n}{b}\log(1 + bQ_T + bV^2) + aVD\log(1 + Q_T/V^2) + D^2\right)$$

Here $Q_T = \min_\mu \sum_{t=1}^T \|v_t - \mu\|$

In the statement of the theorem in [62], they assume that $h_t = h$ for all t and $h'(w^\top v_t) \in [-a, 0]$ for all $w \in \mathcal{D}$. However, they later note that the proof of the theorem is flexible enough to handle different functions h_t for different t. Furthermore, the proof only requires the bound a on the magnitude of the first derivatives at the points w_t, which the algorithm produces, and not the entire domain $\mathcal{D}$.

Theorem 4.6. *For $w \in \Delta$, any sequence of returns $r_1, \ldots, r_T \in \mathbb{R}_+^n$, define $f_t(w) = -\log(r_t^\top w)$. The AdaCurv ONS updates (Equation* (4.5)*) with $\epsilon = 1$ satisfy the regret bound:*

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) = O\left(nR^2\log(1 + Q_T + n) + \sqrt{n}R\log(1 + Q_T/n) + 1\right)$$

Here $Q_T = \min_\mu \sum_{t=1}^T \|r_t - \mu\|_2^2 = \sum_{t=1}^T \|r_t - \bar{r}_T\|_2^2$, where $\bar{r}_T = \frac{1}{T}\sum_{t=1}^T r_t$.

Proof. The iterates of AdaCurvONS are computed as:

$$w_t = \arg\min_{w \in \Delta_n} \frac{1}{2}\|w\|_2^2 + \sum_{s=1}^{t-1} \left(f_s(w_s) - \frac{r_s^\top (w - w_s)}{(r_s^\top w_s)^\top} + \frac{(r_s^\top (w - w_s))^2}{2(r_s^\top w_s)} \right)$$

We can replace r_t with $\tilde{r}_t = \frac{r_t}{\min_i r_t(i)}$ in the above equation without changing the iterates as the optimization is invariant to scaling. We can apply Theorem B.8 with $v_t = \tilde{r}_t$. The function $h_t(x) = f_t(w_t) - \frac{x - \tilde{r}_s^\top w_s}{(\tilde{r}_s^\top w_s)^\top} + \frac{(x - \tilde{r}_s^\top w_s)^2}{2(\tilde{r}_s^\top w_s)}$. This gives $h_t'(\tilde{r}_t^\top w_t) = \frac{-1}{\tilde{r}_t^\top w_t} \in [-R, 0]$ and $h_t''(\tilde{r}_t^\top w) = \frac{1}{\tilde{r}_t^\top w_t} \geq 1$. Thus we have $\|\tilde{r}_t\| \leq \sqrt{n} = V$, $D = 1$, $a = R$ and $b = 1$. So, we have the regret bound:

$$\sum_{t=1}^{T} f_t(w_t) - f_t(w) = O\left(nR^2 \log(1 + Q_T + n) + \sqrt{n} R \log(1 + Q_T/n) + 1 \right)$$

Here $Q_T = \min_\mu \sum_{t=1}^T \|r_t - \mu\| = \sum_{t=1}^T \|r_t - \bar{r}_T\|$, where $\bar{r}_T = \frac{1}{T}\sum_{t=1}^T r_t$. ■