# Online Submodular Maximization via Online Convex Optimization

**Alireza Kazemipour**
kazemipo@ualberta.ca

## Abstract

We will: (i) replace the optimistic online mirror ascent[1] algorithm in our target paper, which is old, with the most updated version of this algorithm and prove that the new algorithm's regret matches the regret of the algorithm in the paper, (ii) give an algorithm based on the idea of adaptive learning for dynamic environments that matches the dynamic regret of the optimistic online mirror ascent in the target paper, but without the need of knowing the path-length in advance which is a violation of the adversarial input assumption happening in our target paper.

## Clarification

Throughout this report by mentioning the phrase **target paper**, we mean the paper that we were supposed to summarize and work on.

## 1 Setup

The problem of interest is online submodular maximization (OSM) via online convex optimization (OCO). In this setting inputs $\mathcal{X} \subseteq \{0,1\}^n$ represent general matroids and the reward function $f : \mathcal{X} \to \mathbb{R}_{\geq 0} \in \mathcal{F}$ is monotone and submodular. In order to turn this problem into an OCO instant, the integral input should be transformed to a fractional counterpart that is convex and compact and be transformed back. By defining $\mathcal{Y} \coloneqq \text{conv}(\mathcal{X})$ to be the convex hull of $\mathcal{X}$ and having a randomized mapping $\Xi : \mathcal{Y} \to \mathcal{X}$ to map a fractional solution $\mathbf{y} \in \mathcal{Y}$ and, possibly, a source of randomness back to an integral solution $\mathbf{x} \in \mathcal{X}$, these requirements are met. Also in the OCO the reward function being optimized should be $L$-Lipschitz concave w.r.t to a norm over its input for some $L \in \mathbb{R}_{\geq 0}$ but in order to reduce an OSM problem to an OCO counterpart, more restrictions should be put on the reward function evaluating fractional solutions $\mathcal{Y}$. The following assumption sums up the necessary properties of such reward functions:

**Assumption** (Sandwich Property)**.** There exists an $\alpha \in (0,1]$, and $L \in \mathbb{R}_{\geq 0}$, and a randomized rounding $\Xi : \mathcal{Y} \to \mathcal{X}$ such that, for every $f :\in \mathcal{X} \to \mathbb{R}_{\geq 0} \in \mathcal{F}$ there exists a concave function $\tilde{f} : \mathcal{Y} \to \mathbb{R}$ that is not only $L$-Lipschitz but also:

$$\tilde{f}(\mathbf{x}) \geq f(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$$

and

$$\mathbb{E}_{\Xi}[f(\Xi(\mathbf{y}))] \geq \alpha \cdot \tilde{f}(\mathbf{y}), \forall \mathbf{y} \in \mathcal{Y}$$

## 2 Using an instantiation of OCO to solve OSM

Having introduced the problem in Section 1, assuming a randomized rounding $\Xi$ and a concave relaxed function $\tilde{f}$ exist, so integral to fractional conversion and vice-versa is done straightforwardly,

[1] In the literature it is more common to talk about online mirror descent but since we are dealing with a maximization problem, it turns into online mirror *ascent*.

in this section we investigate online mirror ascent (OMA) as an OCO method, and one of its variants, optimistic online mirror ascent used in the target paper. The setting is as follows: inputs are in full-information adversarially chosen configuration, the performance measure is the dynamic regret and the problem could be in the optimistic setting and/or not.

### 2.1 Optimistic Online Mirror Ascent: Two-Update-Per-Step

Let the mirror map $\Phi : \mathcal{X} \to \mathbb{R}$ be strongly (thus strictly) convex, smooth and twice differentiable and $B_\Phi$ denote the Bregman divergence with respect to $\Phi$. Then the following algorithm represents the OMA applied on the fractional solutions of an OSM instance:

**Algorithm 1:** Online Mirror Ascent

**Require:** $\eta \in \mathbb{R}_{\geq 0}$ `/* learning rate */`
$\Phi : \mathcal{Y} \to \mathbb{R}$ `/* mirror map */`
Let $\mathbf{y}_1 = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \Phi(\mathbf{y})$
**for** $t = 1$ to $T$ **do**
  Play $\mathbf{y}_t$.
  Observe the reward function $\tilde{f}_t$ and let $\nabla_t = \nabla \tilde{f}_t(\mathbf{y}_t)$
  $\mathbf{y}_{t+1} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \langle \mathbf{y}, \nabla_t \rangle - \frac{B_\Phi(\mathbf{y}; \mathbf{y}_t)}{\eta}$
**end for**

The Algorithm 1 is proved to enjoy sublinear regret under the worst-case analysis [1, 3]; however, Rakhlin and Sridharan [9] later introduced the optimistic online mirror ascent (OOMA) as a variant that takes advantage of benign inputs that do not represent the worst-case scenario. So some predictions about the input in OOMA can be made, and those predictions are assumed to be available to the learner in advance. As a consequence obtaining zero regret is possible in this setting. Moreover OOMA is robust enough such that if no prediction could be made, or the predictions were wrong, or the learner was facing the worst-case scenario, it can default back to OMA and still enjoys a sublinear regret.

Let $\Phi$ and $B_\Phi$ be the same as they were in OMA. Now let $M_t$ denote a prediction about $\tilde{f}_t$. OOMA, as was introduced in Rakhlin and Sridharan [9] and was used in our target paper, takes the following form to take advantage of such predictions:

**Algorithm 2:** Optimistic Online Mirror Ascent: Two-Update-Per-Step

**Require:** $\eta \in \mathbb{R}_{\geq 0}$ `/* learning rate */`
$\Phi : \mathcal{Y} \to \mathbb{R}$ `/* mirror map */`
$M_2, M_3, \cdots, M_{T+1}$ `/* Sequence of predictions */`
Let $\mathbf{y}_1 = \mathbf{z}_1 = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \Phi(\mathbf{y})$
**for** $t = 1$ to $T$ **do**
  Play $\mathbf{y}_t$.
  Observe the reward function $\tilde{f}_t$ and let $\nabla_t = \nabla \tilde{f}_t(\mathbf{y}_t)$
  $\mathbf{z}_{t+1} = \operatorname{argmax}_{\mathbf{z} \in \mathcal{Y}} \langle \mathbf{z}, \nabla_t \rangle - \frac{B_\Phi(\mathbf{z}; \mathbf{z}_t)}{\eta}$ `/* Adapt the secondary decision */`
  $\mathbf{y}_{t+1} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \langle \mathbf{y}, M_{t+1} \rangle - \frac{B_\Phi(\mathbf{y}; \mathbf{z}_{t+1})}{\eta}$ `/* Adapt the primary decision */`
**end for**

OOMA as in Algorithm 2 includes a two-update-per-step procedure. Initially an intermediate decision called the secondary decision is computed the same way the decision is obtain in OMA; but, instead of using this decision, the learner computes another decision called the primary decision according to the precision of the prediction $M_{t+1}$ of the future reward $\tilde{f}_{t+1}$ and the secondary decision that was just computed. Note that if $(M_t)_{t=2}^{T+1} = 0$, then OOMA would be equal to OMA and if $(M_t = \nabla_{t-1})_{t=2}^{T+1}$, we will show that static/dynamic regret would be zero.

### 2.2 Regret Bounds

Before starting the analysis, it is useful to revisit definitions of the static and the dynamic regrets. Let $\pi_{\mathbf{w}}$ be the policy that the learner follows when it chooses actions $\mathbf{w} \in \mathcal{W}$ where $\mathcal{W}$ is some domain, let $T$ be the horizon of the interaction, and let the problem be of the form of an OSM with adversarial inputs.

**Static** regret (performance against the best fixed decision in hindsight) is defined as:

$$SR_T(\pi_{\mathbf{x}}) \coloneqq \sup_{(f_t)_{t=1}^T \in \mathcal{F}^T} \left\{ \max_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x}) - \mathbb{E}\left[ \sum_{t=1}^T f_t(\mathbf{x}_t) \right] \right\}$$

where the expectation is with respect to the randomness present in the interaction.

**Dynamic** regret (performance against the best-per-step sequence of decisions) is defined as:

$$DR_T(\pi_{\mathbf{x}}) \coloneqq \sup_{(f_t)_{t=1}^T \in \mathcal{F}^T} \left\{ \sum_{t=1}^T \max_{\mathbf{u}_t \in \mathcal{X}} f_t(\mathbf{u}_t) - \mathbb{E}\left[ \sum_{t=1}^T f_t(\mathbf{x}_t) \right] \right\}$$

If $\mathbf{u}_t$s change rapidly and radically from time to time, then it becomes notoriously hard to expect the learner learn something meaningful as in the extreme case $\mathbf{u}_t$s would change like a noise. Hence usually a limit is imposed on how much the adversary can change the comparator sequence $(\mathbf{u}_t)_{t=1}^T$. One such limit is known as the *path length*. A path length of $P_T \in \mathbb{R}_{\geq 0}$ over a time horizon $T$ expresses that $\sum_{t=1}^{T-1} \|\mathbf{u}_{t+1} - \mathbf{u}_t\| \leq P_T$. Hence we denote the dynamic regret for a sequence with the path length $P_T$ for the policy $\pi_{\mathbf{y}}$ that has actions $\mathbf{y} \in \mathcal{Y}$, by $DR_T(\pi_{\mathbf{y}}, P_T)$.

**Theorem 1.** *Consider an OSM instance with the concave $L-$Lipschitz relaxed functions $\tilde{f} \in \tilde{\mathcal{F}}$ where the set $\tilde{\mathcal{F}}$ is chosen by the adversary and let the fractional decision set be $\mathcal{Y} \coloneqq conv(\mathcal{X})$. Algorithm 2 configured with a $\rho$-strongly convex, smooth and twice differentiable mirror map $\Phi : \mathcal{Y} \to \mathbb{R}$, where $\rho \in \mathbb{R}_{>0}$; $\|\nabla \Phi(\mathbf{y})\|_* \leq G_\Phi$ for all $\mathbf{y}$; and $D_\Phi \coloneqq \sqrt{\max_{\mathbf{z},\mathbf{y} \in \mathcal{Y}} \{\Phi(\mathbf{z}) - \Phi(\mathbf{y})\}}$ as the diameter of $\mathcal{Y}$ relative to $\Phi$, enjoys the following upper bound on the dynamic regret with the path length $P_T$ over the horizon $T$:*

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \frac{\eta}{2\rho} \sum_{t=1}^T \|\nabla_t - M_t\|_*^2 + \frac{(D_\Phi^2 + 2G_\Phi P_T)}{\eta} \tag{1}$$

*And when the optimal learning rate $\eta^* = \sqrt{\frac{2\rho(D_\Phi^2 + 2G_\Phi P_T)}{\sum_{t=1}^T \|\nabla_t - M_t\|_*^2}}$ is selected, the following holds:*

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \sqrt{\frac{2(D_\Phi^2 + 2G_\Phi P_T)}{\rho} \sum_{t=1}^T \|\nabla_t - M_t\|_*^2} \tag{2}$$

The proof can be found in the target paper.

It is useful to point out: if $M_t = 0$ the bound holds for the setting with no predictions available, if $M_t = \nabla_t$ obtaining zero regret is possible and if $P_T = 0$ we the get the static regret bound.

### 2.3 Optimistic Online Mirror Ascent: One-Update-Per-Step (Our Contribution 1)

Algorithm 2 was *the first* version of OOMA but not *the last*, nor *the best*. The target paper did not account for the latest result in this area which is due to Joulani et al. [4] where they introduce a variant of OOMA that only requires one update per step, and still matches the static regret bound of OOMA proved by Rakhlin and Sridharan [9, 10].

We, with the help of Orabona [7], use the idea of Joulani et al. [4] and give Algorithm 3 that computes only a single update. However we extend the result of Joulani et al. [4] and prove the **dynamic** regret bounds that match the bounds of Eq. 1 and 2.

**Theorem 2.** *Algorithm 3 under the same conditions stated in Theorem 1, enjoys the regret bounds of Eq. 1 and 2.*

**Algorithm 3:** Optimistic Online Mirror Ascent: One-Update-Per-Step

**Require:** $\eta \in \mathbb{R}_{\geq 0}$ /* learning rate */
$\Phi : \mathcal{Y} \to \mathbb{R}$ /* mirror map */
$M_2, M_3, \cdots, M_{T+1}$ /* Sequence of predictions */
1: Let $\mathbf{y}_1 = \text{argmax}_{\mathbf{y} \in \mathcal{Y}} \Phi(\mathbf{y})$
2: **for** $t = 1$ to $T$ **do**
3: Play $\mathbf{y}_t$.
4: Observe the reward function $\tilde{f}_t$ and let $\nabla_t = \nabla \tilde{f}_t(\mathbf{y}_t)$
5: $\mathbf{y}_{t+1} = \text{argmax}_{\mathbf{y} \in \mathcal{Y}} \langle \mathbf{y}, \nabla_t - M_t + M_{t+1} \rangle - \frac{B_\Phi(\mathbf{y}; \mathbf{y}_t)}{\eta}$
6: **end for**

In order to prove Theorem 3 we separately establish two supplementary lemmas to prove the dynamic regret. They are all due to Orabona [7, 8, 6] with the following differences: (i) Our setting is a maximization problem so we work on online mirror ascent rather than descent (ii) Dynamic regret in non-optimistic setting and static regret in optimistic setting were given but, no proof was given for *the dynamic regret in optimistic setting* (their combination) which is our contribution.

It is useful to note that we have avoided restating assumptions that have already been given in previous sections for convenience.

**Lemma 1.** *The following inequality holds for OMA as in Algorithm 1:*

$$\eta(\tilde{f}_t(\mathbf{u}) - \tilde{f}_t(\mathbf{y}_t)) \leq \langle \eta \nabla_t, \mathbf{u} - \mathbf{y}_t \rangle \leq B_\Phi(\mathbf{u}; \mathbf{y}_t) - B_\Phi(\mathbf{u}; \mathbf{y}_{t+1}) + \frac{\eta^2}{2\rho} \|\nabla_t\|_*^2, \forall \mathbf{u} \in \mathcal{Y}$$

*Proof.* From the optimality condition for the update rule of Algorithm 1 (Line 5) and the KKT theorem [Hazan [2], Theorem 2.2] we have:

$$\langle \eta \nabla_t - \nabla_\Phi(\mathbf{y}_{t+1}) + \nabla_\Phi(\mathbf{y}_t), \mathbf{y}_{t+1} - \mathbf{u} \rangle \geq 0, \forall \mathbf{u} \in \mathcal{Y} \tag{3}$$

Also form the definition of supergradients we have:

$$\begin{aligned} \langle \eta \nabla_t, \mathbf{u} - \mathbf{y}_t \rangle &= \langle \nabla_\Phi(\mathbf{y}_{t+1}) - \nabla_\Phi(\mathbf{y}_t) - \eta \nabla_t, \mathbf{y}_{t+1} - \mathbf{u} \rangle \\ &\quad + \langle \nabla_\Phi(\mathbf{y}_t) - \nabla_\Phi(\mathbf{y}_{t+1}), \mathbf{y}_{t+1} - \mathbf{u} \rangle + \langle \eta \nabla_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle \end{aligned} \tag{4}$$

Thus using Eq. 3

$$\langle \eta \nabla_t, \mathbf{u} - \mathbf{y}_t \rangle \leq \langle \nabla_\Phi(\mathbf{y}_t) - \nabla_\Phi(\mathbf{y}_{t+1}), \mathbf{y}_{t+1} - \mathbf{u} \rangle + \langle \eta \nabla_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle \tag{5}$$

The following is according to the definition of Bregman divergence [2]. For any $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in the same convex set that constitutes the domain of $\Phi$:

$$(\mathbf{a} - \mathbf{b})^\top (\nabla \Phi(\mathbf{c}) - \nabla \Phi(\mathbf{b})) = B_\Phi(\mathbf{a}; \mathbf{b}) - B_\Phi(\mathbf{a}; \mathbf{c}) + B_\Phi(\mathbf{b}; \mathbf{c})$$

Hence

$$\langle \eta \nabla_t, \mathbf{u} - \mathbf{y}_t \rangle \leq B_\Phi(\mathbf{u}; \mathbf{y}_t) - B_\Phi(\mathbf{u}; \mathbf{y}_{t+1}) - B_\Phi(\mathbf{y}_{t+1}; \mathbf{y}_t) + \langle \eta \nabla_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle \tag{6}$$

Also, for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \inf_{\lambda \in \mathbb{R}_{\geq 0}} \left\{ \frac{1}{2\lambda} \|\mathbf{a}\|_*^2 + \frac{\lambda}{2} \|\mathbf{b}\|^2 \right\}$$

Thus by choosing $\lambda = \rho$:

$$\langle \eta \nabla_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle \leq \frac{\eta^2}{2\rho} \|\nabla_t\|_*^2 + \frac{\rho}{2} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 \tag{7}$$

Plugging Eq. 7 into Eq. 6:

$$\begin{aligned} \langle \eta \nabla_t, \mathbf{u} - \mathbf{y}_t \rangle &\leq B_\Phi(\mathbf{u}; \mathbf{y}_t) - B_\Phi(\mathbf{u}; \mathbf{y}_{t+1}) - B_\Phi(\mathbf{y}_{t+1}; \mathbf{y}_t) + \langle \eta \nabla_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle \\ &\leq B_\Phi(\mathbf{u}; \mathbf{y}_t) - B_\Phi(\mathbf{u}; \mathbf{y}_{t+1}) - B_\Phi(\mathbf{y}_{t+1}; \mathbf{y}_t) + \frac{\eta^2}{2\rho} \|\nabla_t\|_*^2 + \frac{\rho}{2} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 \end{aligned} \tag{8}$$

Since $\Phi$ is $\rho$-strongly convex, then $B_\Phi(\mathbf{y}_{t+1};\mathbf{y}_t) \geq \frac{\rho}{2}\|\mathbf{y}_{t+1}-\mathbf{y}_t\|^2$ hence Eq. 8 is turned into:

$$\begin{aligned}\langle \eta\nabla_t, \mathbf{u}-\mathbf{y}_t\rangle &\leq B_\Phi(\mathbf{u};\mathbf{y}_t) - B_\Phi(\mathbf{u};\mathbf{y}_{t+1}) - B_\Phi(\mathbf{y}_{t+1};\mathbf{y}_t) + \frac{\eta^2}{2\rho}\|\nabla_t\|_*^2 + \frac{\rho}{2}\|\mathbf{y}_{t+1}-\mathbf{y}_t\|^2 \\ &\leq B_\Phi(\mathbf{u};\mathbf{y}_t) - B_\Phi(\mathbf{u};\mathbf{y}_{t+1}) + \frac{\eta^2}{2\rho}\|\nabla_t\|_*^2 \qquad \square\end{aligned}$$

**Lemma 2.** *The dynamic regret of OMA as in Algorithm 1 is the following:*

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \frac{D_\Phi^2 + 2G_\Phi P_T}{\eta} + \sum_{t=1}^T \frac{\eta}{2\rho}\|\nabla_t\|_*^2$$

*Proof.* From Lemma 1 we have:

$$\begin{aligned}\eta(\tilde{f}_t(\mathbf{u}_t) - \tilde{f}_t(\mathbf{y}_t)) \leq \langle \eta\nabla_t, \mathbf{u}_t-\mathbf{y}_t\rangle &\leq B_\Phi(\mathbf{u}_t;\mathbf{y}_t) - B_\Phi(\mathbf{u}_t;\mathbf{y}_{t+1}) + \frac{\eta^2}{2\rho}\|\nabla_t\|_*^2 \\ &\leq B_\Phi(\mathbf{u}_t;\mathbf{y}_t) - B_\Phi(\mathbf{u}_{t+1};\mathbf{y}_{t+1}) + B_\Phi(\mathbf{u}_{t+1};\mathbf{y}_{t+1}) - B_\Phi(\mathbf{u}_t;\mathbf{y}_{t+1}) + \frac{\eta^2}{2\rho}\|\nabla_t\|_*^2\end{aligned}$$

By diving the sides by $\eta$ and summing over time:

$$\begin{aligned}DR_T(\pi_{\mathbf{y}}, P_T) \leq &\underbrace{\sum_{t=1}^T \frac{B_\Phi(\mathbf{u}_t;\mathbf{y}_t) - B_\Phi(\mathbf{u}_{t+1};\mathbf{y}_{t+1})}{\eta}}_{(I)} + \underbrace{\sum_{t=1}^T \frac{B_\Phi(\mathbf{u}_{t+1};\mathbf{y}_{t+1}) - B_\Phi(\mathbf{u}_t;\mathbf{y}_{t+1})}{\eta}}_{(II)} \\ &+ \sum_{t=1}^T \frac{\eta}{2\rho}\|\nabla_t\|_*^2\end{aligned} \tag{9}$$

Let us take care of terms (I) and (II) separately.

(I) is a telescopic sum hence:

$$(I) = \frac{B_\Phi(\mathbf{u}_1;\mathbf{y}_1) - B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_{T+1})}{\eta}$$

For (II) we have:

$$\begin{aligned}(II) &= \frac{1}{\eta}\left[\phi(\mathbf{u}_{t+1}) - \phi(\mathbf{u}_1) + \sum_{t=1}^T \langle \nabla\Phi(\mathbf{y}_{t+1}), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle\right] \\ &= \frac{1}{\eta}\left[\phi(\mathbf{u}_{t+1}) - \phi(\mathbf{u}_1) + \sum_{t=1}^T \langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1) + \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle\right] \\ &= \frac{1}{\eta}\left[\phi(\mathbf{u}_{t+1}) - \phi(\mathbf{u}_1) + \underbrace{\sum_{t=1}^T \langle \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle}_{\text{telescopic sum}} + \sum_{t=1}^T \langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle\right] \\ &= \frac{1}{\eta}\left[\phi(\mathbf{u}_{t+1}) - \phi(\mathbf{u}_1) + \langle \nabla\Phi(\mathbf{y}_1), \mathbf{u}_1 - \mathbf{u}_{T+1}\rangle + \sum_{t=1}^T \langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle\right] \\ &= \frac{1}{\eta}\left[B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_1) - B_\Phi(\mathbf{u}_1;\mathbf{y}_1) + \sum_{t=1}^T \langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle\right]\end{aligned}$$

By plugging (I) and (II) into Eq. 9 we have:

$$\begin{aligned}
DR_T(\pi_{\mathbf{y}}, P_T) &\leq \frac{B_\Phi(\mathbf{u}_1;\mathbf{y}_1) - B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_{T+1})}{\eta} + \frac{B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_1) - B_\Phi(\mathbf{u}_1;\mathbf{y}_1)}{\eta} \\
&+ \frac{1}{\eta}\sum_{t=1}^{T}\langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle + \sum_{t=1}^{T}\frac{\eta}{2\rho}\|\nabla_t\|_*^2 \\
&= \underbrace{\frac{B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_1) - B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_{T+1})}{\eta}}_{(III)} + \frac{1}{\eta}\sum_{t=1}^{T}\underbrace{\langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle}_{(IV)} + \sum_{t=1}^{T}\frac{\eta}{2\rho}\|\nabla_t\|_*^2
\end{aligned}$$

Now let us take care of terms (III) and (IV) separately.

For (III) since the Bregman divergence is always non-negative, we have:

$$\begin{aligned}
(III) &= \frac{B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_1) - B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_{T+1})}{\eta} \\
&\leq \frac{B_\Phi(\mathbf{u}_{T+1};\mathbf{y}_1)}{\eta} \\
&\leq \frac{D_\Phi^2}{\eta}
\end{aligned}$$

For (IV) using Cauchy-Shwarz inequality and that $\|\nabla\Phi(\mathbf{y})\|_* \leq G_\Phi$, we have:

$$\begin{aligned}
(IV) &= \sum_{t=1}^{T}\langle \nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1), \mathbf{u}_t - \mathbf{u}_{t+1}\rangle \\
&\leq \sum_{t=1}^{T}\|\nabla\Phi(\mathbf{y}_{t+1}) - \nabla\Phi(\mathbf{y}_1)\|_*\|\mathbf{u}_t - \mathbf{u}_{t+1}\| \\
&\leq 2G_\Phi P_T
\end{aligned}$$

Putting everything together we have:

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \frac{D_\Phi^2 + 2G_\Phi P_T}{\eta} + \sum_{t=1}^{T}\frac{\eta}{2\rho}\|\nabla_t\|_*^2 \qquad \square$$

Now we have all the tools to prove Theorem 2.

*Proof.* By replacing $\nabla_t$ with $\nabla_t - M_t + M_{t+1}$, and $\mathbf{u}$ with $\mathbf{u}_t$ in Eq. 6 and summing over time, we have:

$$\begin{aligned}
\underbrace{\sum_{t=1}^{T}\langle \nabla_t - M_t + M_{t+1}, \mathbf{u}_t - \mathbf{y}_t\rangle}_{(V)} &\leq \underbrace{\sum_{t=1}^{T}\frac{B_\Phi(\mathbf{u}_t;\mathbf{y}_t) - B_\Phi(\mathbf{u}_t;\mathbf{y}_{t+1}) - B_\Phi(\mathbf{y}_{t+1};\mathbf{y}_t)}{\eta}}_{\square} \\
&+ \underbrace{\sum_{t=1}^{T}\langle \nabla_t - M_t + M_{t+1}, \mathbf{y}_{t+1} - \mathbf{y}_t\rangle}_{(VI)}
\end{aligned}$$

Let us expand (V) and (VI) separately.

For (V) we have:

$$(V) = \underbrace{\sum_{t=1}^{T}\langle \nabla_t, \mathbf{u}_t - \mathbf{y}_t\rangle}_{DR_T(\pi_{\mathbf{y}}, P_T)} + \sum_{t=1}^{T}\langle M_{t+1} - M_t, \mathbf{u}_t\rangle - \sum_{t=1}^{T}\langle M_{t+1} - M_t, \mathbf{y}_t\rangle$$

For (VI) we have:

$$(VI) = \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \sum_{t=1}^{T} \langle M_{t+1}, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle$$

Putting everything together we have:

$$DR_T(\pi_{\mathbf{y}}, P_T) + \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{u}_t \rangle - \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{y}_t \rangle \leq$$

$$\square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \sum_{t=1}^{T} \langle M_{t+1}, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle$$

Which implies:

$$DR_T(\pi_{\mathbf{y}}, P_T) + \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{u}_t \rangle \leq$$

$$\square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \sum_{t=1}^{T} \langle M_{t+1}, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{y}_t \rangle$$

Which implies:

$$DR_T(\pi_{\mathbf{y}}, P_T) + \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{u}_t \rangle \leq \square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \sum_{t=1}^{T} (\langle M_{t+1}, \mathbf{y}_{t+1} \rangle - \langle M_t, \mathbf{y}_t \rangle)$$

$$\leq \square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \langle M_{T+1}, \mathbf{y}_{T+1} \rangle - \langle M_1, \mathbf{y}_1 \rangle$$

Hence

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle + \langle M_{T+1}, \mathbf{y}_{T+1} \rangle - \langle M_1, \mathbf{y}_1 \rangle - \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{u}_t \rangle$$

Given that the left-hand side is independent of $M_{T+1}$, we can safely set it to zero [see Joulani et al. [4], Appendix 6]:

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle - \langle M_1, \mathbf{y}_1 \rangle - \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{u}_t \rangle$$

$M_1$ is the prediction of the supergradients at $t = 1$ and $\mathbf{y}_1$, as the optimal decision, tries to move in the direction of $M_1$ to maximize the objective thus, $\langle M_1, \mathbf{y}_1 \rangle \geq 0$:

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle - \sum_{t=1}^{T} \langle M_{t+1} - M_t, \mathbf{u}_t \rangle$$

Since $M_t$ is prediction about the supergradients of $f_t$ and $\mathbf{u}_t$ maximizes the reward function at $f_t$, thus $M_t \mathbf{u}_t \leq M_{t+1} \mathbf{u}_t, \forall t \in [T]$ (gradient at the optimal point moves the optimal solution the least). Hence:

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \square + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle$$

$$\leq \sum_{t=1}^{T} \frac{B_\Phi(\mathbf{u}_t; \mathbf{y}_t) - B_\Phi(\mathbf{u}_t; \mathbf{y}_{t+1}) - B_\Phi(\mathbf{y}_{t+1}; \mathbf{y}_t)}{\eta} + \sum_{t=1}^{T} \langle \nabla_t - M_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle$$

And similar to the procedure of Lemma 2 with the difference that instead of $\nabla_t$, $\nabla_t - M_t$ appears on the right hand side, we get:

$$DR_T(\pi_{\mathbf{y}}, P_T) \leq \frac{D_\Phi^2 + 2G_\Phi P_T}{\eta} + \sum_{t=1}^{T} \frac{\eta}{2\rho} \|\nabla_t - M_t\|_*^2 \qquad \square$$

The optimal learning rate of Eq. 2 is found by taking the derivative of the right hand side w.r.t $\eta$.

## 3 Optimal Learning Rate: Without the Prior Knowledge of Path Length (Our Contribution 2)

The optimal learning rate of Theorem 1 is obtained when the learner knows the path length $P_T$; this is a violation of the assumption on the inputs being adversarial as the learner knows how much the adversary can change the comparator sequence. Zhang et al. [12] has given an algorithm for online gradient ascent (OGA) which does not need the path length to be known in advance. We, with the help of Orabona [8] who explained the the same algorithm of Zhang et al. [12], give Algorithm 4 which is for **OOMA** (as our contribution) when $\Phi = \frac{1}{2}\|\mathbf{y}\|_2^2$ which is a common choice for the mirror map [2, 5].

The optimal learning rate used to be $\eta^* = \sqrt{\frac{2\rho(D_\Phi^2 + 2G_\Phi P_T)}{\sum_{t=1}^T \|\nabla_t - M_t\|_*^2}}$, by assumption that $f_t$ is $L$-Lipschitz we have $\|\nabla_t\|_* \leq L$ and since $M_t$ is the prediction about the supergradients of $f_t$, it is reasonable to also assume that $\|M_t\|_* \leq L$. Also since $\sum_{t=1}^{T-1} \|\mathbf{u}_{t+1} - \mathbf{u}_t\| \leq P_T$, then $P_T \leq D_\Phi T$ and when $\Phi = \frac{1}{2}\|\mathbf{y}\|_2^2$, then $D_\Phi = G_\Phi$. Hence by replacing the minimum and maximum of the parameters, $\eta^*$ is in the following range:

$$\frac{\sqrt{2\rho}D_\Phi}{2L\sqrt{T}} \leq \eta^* \leq \frac{\sqrt{2\rho}D_\Phi}{2L\sqrt{T}} \cdot \sqrt{1+2T}$$

So consider a grid of $N = 1 + \left\lceil \log_2 \sqrt{4+8T} \right\rceil$ learning rates where $\eta^{(i)} = \frac{\sqrt{2\rho}D_\Phi 2^{i-1}}{2L\sqrt{T}}$ for $i = 1, \cdots, N$ which results in $\eta^{(1)} = \frac{\sqrt{2\rho}D_\Phi}{2L\sqrt{T}}$ and $\eta^{(N)} = \frac{\sqrt{2\rho}D_\Phi}{2L\sqrt{T}} \cdot \sqrt{1+2T}$. This grid implies that there exists an index $i^*$:

$$\eta^{(i^*)} \leq \eta^* \leq 2\eta^{(i^*)} \tag{10}$$

Thus leveraging the idea of *exponentially weighted average forecaster* (EWA) [1], we can run multiple instances of OOMA with different learning rates that belong to the aforementioned grid (where Eq. 10 guarantees that at least one of them is the optimal), and treat each of these OOMA instances as an expert. EWA ensures that the best expert will be eventually followed which is translated to finding the optimal learning rate in our problem.

**Algorithm 4:** Optimistic Online Mirror Ascent: One-Update-Per-Step with Adaptive Learning Rate

**Require:** $\alpha \in \mathbb{R}_{\geq 0}$ `/* learning rate */`
$\Phi : \mathcal{Y} \to \mathbb{R}$ `/* mirror map */`
$M_2, M_3, \cdots, M_{T+1}$ `/* Sequence of predictions */`

Let $N = 1 + \left\lceil \log_2 \sqrt{4+8T} \right\rceil$
Let $\eta^{(i)} = \frac{\sqrt{2\rho}D_\Phi 2^{i-1}}{2L\sqrt{T}}$, $i = 1, \cdots, N$
Let $\mathbf{p}_1 = [\frac{1}{N}, \cdots, \frac{1}{N}] \in [0,1]^N$ `/* Uniform prior over experts */`
Let $\mathbf{y}_1 = \operatorname{argmax}_{\mathbf{y}\in\mathcal{Y}} \Phi(\mathbf{y})$
**for** $t = 1$ to $T$ **do**
  Play $\mathbf{y}_t$.
  Observe the reward function $\tilde{f}_t$ and let $\nabla_t = \nabla \tilde{f}_t(\mathbf{y}_t)$
  Update each expert: $\mathbf{y}_{t+1}^{(i)} = \operatorname{argmax}_{\mathbf{y}\in\mathcal{Y}} \langle \mathbf{y}, \nabla_t - M_t + M_{t+1} \rangle - \frac{B_\Phi(\mathbf{y};\mathbf{y}_t)}{\eta^{(i)}}$, $i = 1, \cdots, N$
  Update each weight: $\mathbf{p}_{t+1}^{(i)} = \frac{\mathbf{p}_t^{(i)} \exp\left(\alpha \tilde{f}_t\left(\mathbf{y}_t^{(i)}\right)\right)}{\sum_{j=1}^N \mathbf{p}_t^{(j)} \exp\left(\alpha \tilde{f}_t\left(\mathbf{y}_t^{(j)}\right)\right)}$, $i = 1, \cdots, N$
  $\mathbf{y}_{t+1} = \sum_{i=1}^N \mathbf{p}_{t+1}^{(i)} \mathbf{y}_{t+1}^{(i)}$
**end for**

For proving the regret bound of Algorithm 4, it is necessary to state the following lemma of Zhang et al. [12] which is based on EWA in Cesa-Bianchi and Lugosi [1].

**Lemma 3.** *Let $\mathbf{p}_1^{(i)}$ denote the initial probability of following the expert $(i)$ in EWA, $\alpha \in \mathbb{R}_{\geq 0}$, and, without loss of generality, assume $\tilde{f}_t(\cdot)$ is bounded* [2] *in $[0, c]$ at all time steps. The following is the regret of expert $(i)$ in Algorithm 4:*

$$\max_i \left( \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) - \frac{1}{\alpha} \ln \frac{1}{\mathbf{p}_1^{(i)}} \right) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t) \leq \frac{\alpha T c^2}{8}$$

*Then by choosing $\alpha = \sqrt{\frac{8 \ln N}{Tc^2}}$ to minimize the upper bound and the fact that $\mathbf{p}_1^{(i)} = \frac{1}{N}$, for all $i = 1, \cdots, N$ we have:*

$$\sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t) \leq \frac{c\sqrt{2T \ln N}}{2}$$

*Note that if the upper bound is valid for the* max *member, then it is true for all members.*

*Proof.* Let:

$$F_t^{(i)} = \sum_{\tau=1}^{t} \tilde{f}_\tau \left( \mathbf{y}_\tau^{(i)} \right), W_t = \sum_{i=1}^{N} \mathbf{p}_1^{(i)} \exp \left( \alpha F_t^{(i)} \right)$$

Line 9 in Algorithm 4 implies that:

$$\mathbf{p}_t^{(i)} = \frac{\mathbf{p}_1^{(i)} \exp \left( \alpha F_{t-1}^{(i)} \right)}{\sum_{j=1}^{N} \mathbf{p}_1^{(j)} \exp \left( \alpha F_{t-1}^{(j)} \right)}, t \geq 2 \tag{11}$$

Also:

$$\begin{aligned} \ln W_T &= \ln \left( \sum_{i=1}^{N} \mathbf{p}_1^{(i)} \exp \left( \alpha F_T^{(i)} \right) \right) \\ &\geq \ln \left( \max_i \mathbf{p}_1^{(i)} \exp \left( \alpha F_T^{(i)} \right) \right) \\ &= \alpha \max_i \left( F_T^{(i)} - \frac{1}{\alpha} \ln \frac{1}{\mathbf{p}_1^{(i)}} \right) \end{aligned} \tag{12}$$

Next let us bound $\ln \left( \frac{W_t}{W_{t-1}} \right)$. For $t \geq 2$, we have:

$$\begin{aligned} \ln \left( \frac{W_t}{W_{t-1}} \right) &= \ln \left( \frac{\sum_{i=1}^{N} \mathbf{p}_1^{(i)} \exp \left( \alpha F_t^{(i)} \right)}{\sum_{j=1}^{N} \mathbf{p}_1^{(j)} \exp \left( \alpha F_{t-1}^{(j)} \right)} \right) \\ &= \ln \left( \frac{\sum_{i=1}^{N} \mathbf{p}_1^{(i)} \exp \left( \alpha F_{t-1}^{(i)} \right) \exp \left( \alpha \tilde{f}_t(\mathbf{y}_t^{(i)}) \right)}{\sum_{j=1}^{N} \mathbf{p}_1^{(j)} \exp \left( \alpha F_{t-1}^{(j)} \right)} \right) \\ &= \ln \left( \sum_i \mathbf{p}_t^{(i)} \exp \left( \alpha \tilde{f}_t \left( \mathbf{y}_t^{(i)} \right) \right) \right) \qquad \text{(Using Eq. 11)} \end{aligned}$$

When $t = 1$, we have:

$$\ln W_1 = \ln \left( \sum_i \mathbf{p}_1^{(i)} \exp \left( \alpha \tilde{f}_1 \left( \mathbf{y}_1^{(i)} \right) \right) \right)$$

Hence

$$\ln W_T = \ln W_1 + \sum_{t=2}^{T} \ln \left( \frac{W_t}{W_{t+1}} \right) = \sum_{t=1}^{T} \ln \left( \sum_i \mathbf{p}_t^{(i)} \exp \left( \alpha \tilde{f}_t \left( \mathbf{y}_t^{(i)} \right) \right) \right) \tag{13}$$

[2] Bounded reward functions is a common and necessary condition in online learning [Lattimore and Szepesvári [5], Chapter 1 and also Szepesvári [11], Chapter 1].

On the other hand, Hoeffding's inequality states that for any random variable $X$ such that $a \leq X \leq b$, and any $s \in \mathbb{R}$:

$$\ln\left(\mathbb{E}\left[\exp(sX)\right]\right) \leq s\mathbb{E}\left[X\right] + \frac{s^2(b-a)^2}{8}$$

So since $0 \leq \tilde{f}_t(\cdot) \leq c$ for all $t$, we apply the Hoeffding's inequality to the inner summand of Eq. 13

$$\begin{aligned}
\ln\left(\sum_i \mathbf{p}_t^{(i)} \exp\left(\alpha \tilde{f}_t\left(\mathbf{y}_t^{(i)}\right)\right)\right) &\leq \alpha \sum_i \mathbf{p}_t^{(i)} \tilde{f}_t\left(\mathbf{y}_t^{(i)}\right) + \frac{\alpha^2 c^2}{8} \\
&\leq \alpha \tilde{f}_t\left(\sum_i \mathbf{p}_t^{(i)} \mathbf{y}_t^{(i)}\right) + \frac{\alpha^2 c^2}{8} \quad \text{(Jensen's inequality for concave functions)} \\
&= \alpha \tilde{f}_t(\mathbf{y}_t) + \frac{\alpha^2 c^2}{8}
\end{aligned}$$

Substituting the above inequality in Eq. 13, we have:

$$\ln W_T \leq \alpha \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t) + \frac{T\alpha^2 c^2}{8}$$

Substituting the above inequality in Eq. 12 and substituting $F_T^{(i)}$ with its explicit value, we have:

$$\max_i \left(\sum_{t=1}^{T} \tilde{f}_t\left(\mathbf{y}_t^{(i)}\right) - \frac{1}{\alpha} \ln \frac{1}{\mathbf{p}_1^{(i)}}\right) \leq \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t) + \frac{T\alpha c^2}{8} \qquad \square$$

**Theorem 3.** *Algorithm 4 matches the optimal regret bound of Theorem 2 up to some polylogarithmic factors without the need of knowing $P_T$ in advance.*

*Proof.* According to Theorem 2, the regret of expert $(i)$ in Algorithm 4 is:

$$\begin{aligned}
&\sum_{t=1}^{T} \tilde{f}_t(\mathbf{u}_t) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) \\
&\leq \frac{\eta^{(i)}}{2\rho} \sum_{t=1}^{T} \|\nabla_t - M_t\|_*^2 + \frac{(D_\Phi^2 + 2G_\Phi P_T)}{\eta^{(i)}}
\end{aligned}$$

And according to Eq.10:

$$\begin{aligned}
&\sum_{t=1}^{T} \tilde{f}_t(\mathbf{u}_t) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) \\
&\leq \frac{\eta^{(i)}}{2\rho} \sum_{t=1}^{T} \|\nabla_t - M_t\|_*^2 + \frac{(D_\Phi^2 + 2G_\Phi P_T)}{\eta^{(i)}} \\
&\leq \frac{\eta^{*}}{2\rho} \sum_{t=1}^{T} \|\nabla_t - M_t\|_*^2 + \frac{2(D_\Phi^2 + 2G_\Phi P_T)}{\eta^{*}}
\end{aligned}$$

By plugging the optimal learning rate of Theorem 2 for $\eta^*$, we have:

$$\sum_{t=1}^{T} \tilde{f}_t(\mathbf{u}_t) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) \leq 3\sqrt{\frac{(D_\Phi^2 + 2G_\Phi P_T)}{2\rho} \sum_{t=1}^{T} \|\nabla_t - M_t\|_*^2}$$

And using Lemma 3:

$$\begin{aligned}
&\sum_{t=1}^{T} \tilde{f}_t(\mathbf{u}_t) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) + \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t^{(i)}) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t) \\
&= \sum_{t=1}^{T} \tilde{f}_t(\mathbf{u}_t) - \sum_{t=1}^{T} \tilde{f}_t(\mathbf{y}_t) \\
&\leq 3\sqrt{\frac{(D_\Phi^2 + 2G_\Phi P_T)}{2\rho} \sum_{t=1}^{T} \|\nabla_t - M_t\|_*^2} + \frac{c\sqrt{2T \ln N}}{2}, \forall \mathbf{u}_t \in \mathcal{Y} \qquad \square
\end{aligned}$$

## References


[1] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[2] E. Hazan. Introduction to Online Convex Optimization. *Foundations and Trends® in Optimization*, 2016.

[3] E. Hazan and S. Kale. Extracting certainty from uncertainty: regret bounded by variation in costs. *Machine Learning*, 2010.

[4] P. Joulani, A. György, and C. Szepesvári. A Modular Analysis of Adaptive (Non-)Convex Optimization: Optimism, Composite Objectives, and Variational Bounds. In *Proceedings of the 28th International Conference on Algorithmic Learning Theory*, 2017.

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

[6] F. Orabona. Online Mirror Descent II, 2019. URL `https://parameterfree.com/2019/10/01/online-mirror-descent-ii-regret-and-mirror-version/`.

[7] F. Orabona. A simple proof for optimistic OMD, 2022. URL `https://parameterfree.com/2022/09/19/a-simple-proof-for-optimistic-omd/`.

[8] F. Orabona. Dynamic regret and ADER, 2024. URL `https://parameterfree.com/2024/07/01/dynamic-regret-and-ader/`.

[9] A. Rakhlin and K. Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*. PMLR, 2013.

[10] S. Rakhlin and K. Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 2013.

[11] C. Szepesvári. *Algorithms for Reinforcement Learning*. Springer nature, 2022.

[12] L. Zhang, S. Lu, and Z.-H. Zhou. Adaptive online learning in dynamic environments. *Advances in neural information processing systems*, 2018.