CSE599i: Online and Adaptive Machine Learning **Winter 2018**

Lecture 2: Online Learning

Lecturer: Kevin Jamieson *Scribes: Lun Li, Wenchao Yang, Zexuan Zhou*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Online learning is a natural extension of the statistical learning theory that was able to handle problems where data are accessible only in sequential orders. It overcomes the lack of the ability of the standard statistical learning method that deals with problems where it is essential for the algorithm to constantly and dynamically adapt the patterns of the newly available data. Some common examples include the hedge fund returns, where it is essential for the algorithm to adopt data on the market to make instant predictions; or travel time estimation on Google Map, where dynamic predictions are required to be made based on bandit feedback. This script is aiming to serve as a scribe note on the first two lectures of CSE 599 on Online learning, with extensions based on theorems and examples on Bubeck's lecture note [2].

1 Intuition and Motivating Examples

The online learning protocol can be described as follows. Suppose $\mathcal{A}$ is a given set of possible actions.

- Player chooses action $a_t \in \mathcal{A}$.
- An adversary selects $z_t \in \mathcal{Z}$ simultaneously.
- Player suffers loss $\ell(a_t, z_t)$.
- Observe z_t.

The objective is to minimize the cumulative regret R_n:

$$R_n = \sum_{t=1}^{n} \ell(a_t, z_t) - \inf_{a \in \mathcal{A}} \sum_{t=1}^{n} \ell(a, z_t),$$

where $\sum_{t=1}^{n} \ell(a_t, z_t)$ is the cumulative loss of the player, and $\inf_{a \in \mathcal{A}} \sum_{t=1}^{n} \ell(a, z_t)$ is the cumulative loss of best action in hindsight. We also have another type of regret, known as expert regret. Suppose that at every time step t, the player receives a set of expert advice $b_t \in \mathcal{A}^{\mathrm{d}}$, in which the advice of the i-th expert is denoted as $b_t(i)$. The goal for the player is to perform as well as the best expert. So, we define the following regret R_n^E, which compares the cumulative regret of the player with respect to the experts.

$$R_n^E = \sum_{t=1}^{n} \ell(a_t, z_t) - \min_{1 \le i \le d} \sum_{t=1}^{n} \ell(b_t(i), z_t)$$

1.1 Example 1: Weather Forecast

In this example, we firstly apply the definition of expert regret. As defined above, at every time step t, the player aims to predict the temperature based on the expert advice $b_t \in \mathcal{A}^{\mathrm{d}}$, in which the advice of the i-th expert is denoted as $b_t(i)$. Here, the experts reveals their advice to the player before he makes his own choice a_t. The aim of the game is to minimize the cumulative regret:

$$R_n^E = \sum_{t=1}^{n} \ell(a_t, z_t) - \min_{1 \le i \le d} \sum_{t=1}^{n} \ell(b_t(i), z_t)$$

If the player is allowed to take a combination of expert advice as his choice, we can define a new loss function $\bar{\ell}(p_t, (b_t, z_t))$ as follows, where $p_t \in \Delta^d = \{p \in \mathbb{R}_+^d, \sum_{i=1}^d p(i) = 1\}$ is a $(d-1)$-simplex.

$$\bar{\ell}(p_t, (b_t, z_t)) = \ell(p_t^T b_t, z_t)$$

Here the player's choice $a_t = p_t^T b_t$ is a weighted average of the expert advice b_t. Let $e_1, ..., e_d$ be the canonical basis of $\mathbb{R}^d$, which correspond to the vertices of the simplex. Then, $\bar{\ell}(e_i, (b_t, z_t)) = \ell(e_i^T b_t, z_t)$ is the loss for the i-th expert at time t. The cumulative regret is defined as follows. In words, the cumulative regret compares the cumulative loss of a weighted average of expert advice to the cumulative loss of the best performing expert in hindsight.

$$R_n^E = \sum_{t=1}^n \bar{\ell}(p_t, (b_t, z_t)) - \min_{1 \le i \le d} \sum_{t=1}^n \bar{\ell}(e_i, (b_t, z_t))$$

.

1.2 Example 2: Sequential Investment

We assume the player has an initial amount of money W_0. At every time step t, he invests his total capital into d assets whose price relatives are $z_t \in \mathbb{R}_+^d$. The proportions of money he allocates are $a_t \in \mathbb{R}_+^d$. At the end of trading period t, the total amount of money he owns is:

$$W_t = W_{t-1} a_t^T z_t = W_0 \prod_{s=1}^t a_s^T z_s$$

It is natural to consider the ratio of rewards to an adversary. In the following formula, $W_0 \prod_{s=1}^n e_i^T z_s$ corresponds to the total capital when $t = n$ if all money is invested into the i-th asset:

$$R_n = \log(\frac{\max_i W_0 \prod_{s=1}^n e_i^T z_s}{W_0 \prod_{s=1}^n a_s^T z_s}) = \sum_{s=1}^n -\log(a_s^T z_s) - \min_i \sum_{s=1}^n -\log(e_i^T z_s)$$

This motivates the so-called log loss $\ell(a, z) = -\log(a^T z)$. The properties of log-loss is shown in Proposition 1 below. In words, the cumulative regret R_n compares the log-loss of the player's choice to the optimal log-loss of the investment into a single asset from the beginning to end. Bubeck (2011) [2] also mentioned another regret function which needs constantly rebalanced portfolios. That is, $\forall t \ge 1, a_t = a$, which means the player rebalanced his allocation of assets to a at the end of each trading period. Then cumulative regret R_n below compares the log-loss of the player's choice to that of the best constantly rebalanced portfolio a in hindsight.

$$R_n = \log(\frac{\max_{a \in \mathcal{A}} W_0 \prod_{s=1}^n a^T z_s}{W_0 \prod_{s=1}^n a_s^T z_s}) = \sum_{s=1}^n -\log(a_s^T z_s) - \min_{a \in \mathcal{A}} \sum_{s=1}^n -\log(a^T z_s)$$

Definition 1. *Let $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X} \subset \mathbb{R}^d$. Then one says that*

- *f is σ-exp concave ($\sigma > 0$) if $x \mapsto exp(-\sigma f(x))$ is a concave function*
- *f is α-strongly convex if it is subdifferentiable and $\forall x \in \mathcal{X}$,*

$$f(x) - f(y) \le \nabla f(x)^T (x - y) - \frac{\alpha}{2} ||x - y||_2^2, \; \forall y \in \mathcal{X}$$

Proposition 1. *The log-loss $(a, z) \in \mathbb{R}_+^d \times \mathbb{R}_+^d \mapsto -\log(a^T z)$ has the following properties.*

- *It is 1-exp concave.*
- *It takes unbounded values and it has unbounded gradient, even when restricted to the $(d-1)$-simplex.*
- *It is not α-strongly convex, for any $\alpha > 0$.*

2 Exponential Weights

We recall that in online optimization the cumulative regret with the $(d-1)$-simplex $\Delta_d = \{p \in R_+^d, \sum_{i=1}^d p(i) = 1\}$ as the action set is defined as

$$R_n = \sum_{t=1}^n \ell(p_t, z_t) - \inf_{q \in \Delta_d} \sum_{t=1}^n \ell(q, z_t),$$

while the expert regret with canonical basis $e_1, ..., e_d$ of $\mathbb{R}^d$ is defined as

$$R_n^E = \sum_{t=1}^n \ell(p_t, z_t) - \inf_{1 \le i \le d} \sum_{t=1}^n \ell(e_i, z_t)$$

In Section 2.1 we describing the strategy of online learning, namely the exponentially weighted average forecaster. Then we prove an expert regret bound for this strategy in the setting of bounded convex losses in Section 2.2.

2.1 Exponentially weighted average forecaster (Exp strategy)

Suppose one can play on the simplex but wants to perform as well as the best vertex in the simplex (i.e. minimize the expert regret). A way to come up with a strategy is to assign a weight to each vertex on the basis of its past performances, and then take the corresponding convex combination of the vertices as its decision p_t. The weight of a vertex should be a non-increasing function of its past cumulative loss. If we choose the exponential function we obtain the following decision:

$$p_t = \sum_{i=1}^d \frac{w_t(i)}{\sum_{j=1}^d w_t(j)} e_i,$$

where

$$w_t(i) = \exp\Big(-\eta \sum_{s=1}^{t-1} \ell(e_i, z_s)\Big)$$

and $\eta > 0$ is a fixed parameter. If we plug $w_t(i)$ into p_t, we have $\forall i \in 1, ..., d$

$$p_t(i) = \frac{\exp\Big(-\eta \sum_{s=1}^{t-1} \ell(e_i, z_s)\Big)}{\sum_{j=1}^d \exp\Big(-\eta \sum_{s=1}^{t-1} \ell(e_j, z_s)\Big)}$$

Note that

$$\begin{aligned} w_t(i) &= \exp\Big(-\eta \sum_{s=1}^{t-1} \ell(e_i, z_s)\Big) \\ &= \exp\Big(-\eta \sum_{s=1}^{t-2} \ell(e_i, z_s)\Big) \exp(-\eta \ell(e_{t-1}, z_{t-1})) \\ &= w_{t-1}(i) \exp(-\eta \ell(e_{t-1}, z_{t-1})) \end{aligned}$$

Thus the computational complexity of one step of the Exp strategy is of order $O(d)$.

2.2 Bounded Convex Loss and Expert Regret

In this section we give an upper bound on the expert regret of the Exp strategy for **bounded convex losses**. We assume the loss is bounded between 0 and 1, i.e., $\ell(p,z) \in [0,1], \forall (p,z) \in \Delta_d \times \mathcal{Z}$. If we have the loss $\ell(p,z) \in [m,M]$ then we can rescale it to be a rescaled loss $\bar{\ell}(a,z) = \frac{\ell(a,z)-m}{M-m}$ so that $\bar{\ell} \in [0,1]$.
We first provide the definition of convexity and introduce Hoeffding's inequality, which will be used in the proof for expert regret bound later.

Definition 2. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if $\forall x, y \in \mathbb{R}^d$, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \forall \lambda \in [0,1]$*

Lemma 1. *(Jensen's Inequality) Let X be a random variable and f be convex. Then $f(E[X]) \leq E[f(x)]$.*

Proof. We prove Jensen's inequality only for the case where X takes value in a finite set $M = \{x_1, ..., x_k\}$ with probability $\mathbb{P}(X = x_i) = p_i$, for $i = 1, ..., k$ s.t. $\sum_{i=1}^d p_i = 1$. We first consider the case where M contains only two elements. In this case we have the following:

$$\begin{aligned}\mathbb{E}[f(X)] &= p_1 f(x_1) + p_2 f(x_2) \\ &\geq f(p_1 x_1 + p_2 x_2) \quad (f \text{ is convex}) \\ &= f(\mathbb{E}[X])\end{aligned}$$

We will prove Jensen's inequality holds true for finite $M \geq 2$ by induction. Suppose M contains k elements and assume that Jensen's inequality holds for distributions on $k-1$ points. We now have the following where the forth line follows from the induction hypothesis.

$$\begin{aligned}\mathbb{E}[f(X)] &= p_1 f(x_1) + p_2 f(x_2) + \cdots + p_k f(x_k) \\ &= (p_1+p_2)\Big(\frac{p_1}{p_1+p_2} f(x_1) + \frac{p_2}{p_1+p_2} f(x_2)\Big) + p_3 f(x_3) + \cdots + p_k f(x_k) \\ &\geq (p_1+p_2) f\Big(\frac{p_1}{p_1+p_2} x_2 + \frac{p_2}{p_1+p_2} x_2\Big) + p_3 f(x_3) + \cdots + p_k f(x_k) \\ &\geq f\Big((p_1+p_2)\Big(\frac{p_1 x_1}{p_1+p_2} + \frac{p_2 x_2}{p_1+p_2}\Big) + p_3 x_3 + \cdots + p_k x_k\Big) \\ &= f(p_1 x_1 + p_2 x_2 + p_3 x_3 + \cdots + p_k x_k) \\ &= f(\mathbb{E}[X])\end{aligned}$$

□

Lemma 2. *(Hoeffding's Inequality) Let X be a real random variable with $a \leq X \leq b$. Then for any $s \in \mathbb{R}$,*

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

Now we prove the following regret bound:

Theorem 1. *If $p :\mapsto \ell(p,z)$ is **convex** $\forall z$, and $\ell(p,z) \in [0,1]$ (**bounded**), i.e., for any convex loss taking values in $[0,1]$, the Exp strategy satisfies:*

$$\begin{aligned}R_n^E &= \sum_{t=1}^n \ell(p_t, z_t) - \inf_{1 \leq i \leq d} \sum_{t=1}^n \ell(e_i, z_t) \\ &\leq \frac{\log d}{\eta} + \frac{n\eta}{8}\end{aligned}$$

In particular with $\eta = 2\sqrt{\frac{2\log d}{n}}$ it satisfies:

$$R_n^E \leq \sqrt{\frac{n \log d}{2}}.$$

Proof. Let $w_t(i) = \exp\Big(- \eta \sum_{s=1}^{t-1} \ell(e_i, z_s)\Big)$ and $W_t = \sum_{i=1}^{d} w_t(i)$ (by definition $w_1(i) = 1$ and $W_1 = d$). First, note that:

$$\begin{aligned}
\log \frac{W_{n+1}}{W_1} &= \log \Big(\sum_{i=1}^{d} w_{n+1}(i)\Big) - \log d \\
&\geq \log \Big(\max_{1\leq i\leq d} w_{n+1}(i)\Big) - \log d \\
&= -\eta \min_{1\leq i\leq d} \sum_{t=1}^{n} \ell(e_i, z_t) - \log d
\end{aligned}$$

Then we note that $\log \frac{W_{n+1}}{W_1} = \sum_{t=1}^{n} \log \frac{W_{t+1}}{W_t}$, and

$$\log \frac{W_{t+1}}{W_1} = \log \Big(\sum_{i=1}^{d} \frac{w_t(i)}{W_t} \exp(-\eta\ell(e_i, z_t))\Big)$$

Note that if we write $p_t(i) : \mathbb{P}(I = i) = \frac{w_t(i)}{W_t}$ then $\sum_{i=1}^{d} \frac{w_t(i)}{W_t} \exp(-\eta\ell(e_i, z_t))$ can be written as $\mathbb{E}_{I\sim p_t}[\exp(-\eta\ell(e_I, z_t))]$ and hence,

$$\begin{aligned}
\log \frac{W_{t+1}}{W_1} &= \log \Big(\sum_{i=1}^{d} \frac{w_t(i)}{W_t} \exp(-\eta\ell(e_i, z_t))\Big) && (1) \\
&= \log(\mathbb{E}[\exp(-\eta\ell(e_I, z_t))]) \\
&\leq -\eta\mathbb{E}[\ell(e_I, z_t)] + \frac{\eta^2}{8} \text{ (Hoeffding's lemma)} \\
&\leq -\eta\ell(\mathbb{E}[e_I], z_t) + \frac{\eta^2}{8} \text{ (Jensen's inequality)} && (2) \\
&= -\eta\ell(p_t, z_t) + \frac{\eta^2}{8} && (3)
\end{aligned}$$

then

$$\begin{aligned}
\log \frac{W_{n+1}}{W_1} &= \sum_{t=1}^{n} \log \frac{W_{t+1}}{W_t} \\
&\leq \sum_{t=1}^{n} \Big(- \eta\ell(p_t, z_t) + \frac{\eta^2}{8}\Big)
\end{aligned}$$

Now we have the lower bound and upper bound for $\log \frac{W_{n+1}}{W_1}$, putting them together we have

$$- \eta \min_{1\leq i\leq d} \sum_{t=1}^{n} \ell(e_i, z_t) - \log d \leq \sum_{t=1}^{n} \Big(- \eta\ell(p_t, z_t) + \frac{\eta^2}{8}\Big)$$

and after rearranging

$$\begin{aligned}R_n^E &= \sum_{t=1}^n \ell(p_t, z_t) - \inf_{1\le i\le d} \sum_{t=1}^n \ell(e_i, z_t)\\ &\le \frac{\log d}{\eta} + \frac{n\eta}{8}\end{aligned} \tag{4}$$

Plug in $\eta = 2\sqrt{\frac{2\log d}{n}}$ to obtain

$$\begin{aligned}R_n^E &\le \frac{\log d}{2\sqrt{\frac{2\log d}{n}}} + \frac{2n\sqrt{\frac{2\log d}{n}}}{8}\\ &= \sqrt{\frac{n\log d}{2}}\end{aligned}$$

Note that mentioned at the beginning of Section 3 that the log-loss takes unbounded values and has unbounded gradients, and thus we fail to use Theorem 1 and need a more advanced scheme.

□

2.3 Exp-concave loss and expert regret

We recall that in section 1.2 the regret of portfolio example can be shown in log-loss form below, and Proposition 1 shows that the log-loss is 1-exp concave. In this section we consider the exp-concave loss, which is another type of convex loss function, and find an upper bound for the Exp strategy.

$$R_n^E = \log\left(\frac{\max_i W_0 \prod_{s=1}^n e_i^T z_s}{W_0 \prod_{s=1}^n a_s^T z_s}\right) = \sum_{s=1}^n -\log(a_s^T z_s) - \min_i \sum_{s=1}^n -\log(e_i^T z_s)$$

Theorem 2. *For any δ-exp-concave loss, the Exp strategy with parameter $\eta = \delta$ satisfies:*

$$R_n^E \le \frac{\log d}{\delta}$$

Proof. In the previous proof it suffices to replace Hoeffing's lemma followed by Jensen's inequality by a single Jensen's inequality applied to $p \mapsto \exp(-\eta\ell(p,z))$. More specifically, let's continue from the previous proof where we use Hoeffing's lemma in equation (1).

$$\begin{aligned}\log\frac{W_{t+1}}{W_1} &= \log\Big(\sum_{i=1}^d \frac{w_t(i)}{W_t}\exp(-\eta\ell(e_i, z_t))\Big)\\ &= \log(\mathbb{E}[\exp(-\eta\ell(e_I, z_t))])\\ &\le \mathbb{E}\log(\exp(-\eta\ell(e_I, z_t))) \text{ (Jensen's inequality)}\\ &= -\eta\mathbb{E}[\ell(e_I, z_t)]\\ &\le -\eta\ell(\mathbb{E}[e_I], z_t) \text{ (Jensen's inequality)}\\ &= -\eta\ell(p_t, z_t)\end{aligned}$$

then as in previous proof we put together the lower bound and upper bound and we will get

$$-\eta \min_{1\le i\le d} \sum_{t=1}^n \ell(e_i, z_t) - \log d \le \sum_{t=1}^n \Big(-\eta\ell(p_t, z_t)\Big)$$

and after rearranging

$$\begin{aligned} R_n^E &= \sum_{t=1}^{n} \ell(p_t, z_t) - \inf_{1 \le i \le d} \sum_{t=1}^{n} \ell(e_i, z_t) \\ &\le \frac{\log d}{\eta} \\ &= \frac{\log d}{\delta} \text{ (plug in } \eta = \delta) \end{aligned}$$

□

2.4 Lower bound

We have already derived an upper bound for the Exp strategy. We here give the lower bound for expert regret with general convex and bounded losses. Note that this also implies lower bound on other regret.

Theorem 3. *Consider the loss $l : (p, z) \in \Delta_d \times \{0,1\}^d \mapsto p^T z \in [0,1]$. For any strategy, the following holds true:*

$$\sup_{n,d} \sup_{adversary} \frac{R_n}{\sqrt{(n/2) \log d}} \ge 1$$

Proof. We skip the proof here and comment that the regret bound for online bounded linear losses is unimprovable. Interested reader please refer to [2] section 2.4. □

2.5 Anytime strategy

One weakness of Exp strategy is that the optimal parameter η depends on n, which in many applications is unknown. If we look back to Theorem 1, we give example of $\eta = 2\sqrt{\frac{2 \log d}{n}}$. Thus one may wonder if there exists a strategy which admits a regret bound uniformly over time. We here provide a time-varying parameter η_t that easily admits a regret bound uniformly over time.

Theorem 4. *For any convex loss with values in $[0,1]$, the Exp strategy with time-varying parameter $\eta_t = 2\sqrt{\frac{\log d}{t}}$ satisfies $\forall n \ge 1$:*

$$R_n^E \le \sqrt{n \log d}$$

Proof. We skip the proof. Interested readers please refer to [2] section 2.5. Note that for this bound the time-varying parameter η_t has nothing to do with n, which means we only need to know about the dimension of the action space d and the time t. □

2.6 Online finite optimization

Section 2.7 (with the argument for why any deterministic strategy would incur linear regret for finite optimization) should be covered and Theorem 2.7 should be stated, with a proof sketch (i.e., its a direct application of Hoeffding-Azuma which you can state.). In previous sections we could play a convex combination of strategies. In this section, we consider the scenario where we can only play a single pure section. In other words, we consider an action set of finite size: $\mathcal{A} = \{1, ..., d\}$. In this case the convexity assumptions of the previous sections, which is based on a set $\mathcal{X} \in \mathbb{R}^d$, does not hold. We will be focusing on bounded losses, $\ell : \mathcal{A} \times \mathcal{Z} \mapsto [0,1]$, with no further restriction. Under such settings, all the deterministic strategies considered so far will all lead to linear regret in finite optimization. This can be shown as the following: without loss of generality. consider a case with $\mathcal{A} = \mathcal{Z} = \{0,1\}$ with zero-one loss $\ell(a, z) = \mathbf{1}_{a \neq z}$. Then if the player takes a_t, a deterministic function of $(z_1, ..., z_{t-1})$, an adversary who knows your strategy can set $z_t = 1 - a_t$ to make $\ell(a_t, z_t) = 1 \Rightarrow \sum_{t=1}^{n} \ell(a, z_t) = n$ regardless the actions of the player. On the other hand, for any choice of $(z_1, ..., z_{t-1})$, we have $\min_{a \in \{0,1\}} \sum_{t=1}^{n} \ell(a, z_t) \leq \frac{n}{2}$, because the maximum value of $\min_{a \in \{0,1\}} \sum_{t=1}^{n} \ell(a, z_t)$ is attend when exact half of z equals 0 and the other half equals 1. So in other words, for any deterministic strategy with a 0-1 loss, we can choose a set of $z_1, ..., z_n$ so that $\min_{a \in \{0,1\}} \sum_{t=1}^{n} \ell(a, z_t) \leq \frac{n}{2}$. This renders $R_n = \sum_{t=1}^{n} \ell(a_t, z_t) - \min_{a \in \{0,1\}} \sum_{t=1}^{n} \ell(a, z_t) \geq \frac{n}{2}$, which gives the worst case lower bound. One way to improve this is to 'trick your adversary' by adding randomization to the decision strategy. So instead of choosing an action deterministically, the player can choose an action $p_t \in \Delta_d$ based on the past. In that case, the regret R_n becomes a well-defined random variable, and therefore an upper bound on R_n can be found that holds either with high probability or in expectation. We may notice that in online finite optimization the player chooses a point $p_t \in \Delta_d$, which is just online optimization over the simplex. In the finite case, a general bounded loss is equivalent to a bounded linear loss on the simplex. Let's consider the loss

$$\bar{\ell}(p, z) = \sum_{a=1}^{d} p(a) \ell(a, z).$$

Note that this loss is linear and takes values in $[0, 1]$. We now show that a regret bound with respect to this modified loss gives a regret bound for the original game. The proof is a direct application of the following Hoeffding-Azuma's inequality for the martingales.

Theorem 5. *(Hoeffding-Azuma's inequality for martingales) Let $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ be a filtration, and $X_1, ..., X_n$ real random variables such that X_t is F_t-measurable, $\mathbb{E}(X_t | F_{t-1}) = 0$ and $X_t \in [A_t, A_t + c_t]$ where A_t is a random variable F_{t-1}-measurable and c_t is a positive constant. Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}(\sum_{t=1}^{n} X_t \geq \epsilon) \leq \exp(-\frac{2\epsilon^2}{\sum_{t=1}^{n} c_t^2}),$$

or equivalently for any $\delta > 0$, with probability at least $1 - \delta$, we have

$$\sum_{t=1}^{n} X_t \leq \sqrt{\frac{\log(\delta^{-1})}{2} \sum_{t=1}^{n} c_t^2}$$

Proof. Interested readers may find the proof of this theorem in the first part of preliminary lemmas 1 Azuma's original paper [1]. □

Theorem 5 will be used to prove the following lemma

Lemma 3. *With probability at least $1 - \delta$ the following holds true:*

$$\sum_{t=1}^{n} \ell(a_t, z_t) - \min_{a \in \mathcal{A}} \sum_{t=1}^{n} \ell(a, z_t) \leq \sum_{t=1}^{n} \bar{\ell}(p_t, z_t) - \min_{q \in \Delta_d} \sum_{t=1}^{n} \bar{\ell}(q, z_t) + \sqrt{\frac{n \log(\delta^{-1})}{2}}$$

Proof. This lemma can be get by directly applying Theorem 5, taking a filtration $\mathcal{F}_t = \sigma(a_1, ..., a_t)$ and $X_t = \ell(a_t, z_t) - \sum_{a=1}^{d} p_t(a) \ell(a, z_t)$. □

We now apply the Exp strategy of Section 2.1 to the modified loss $\bar{\ell}(p, z)$. Recall that this strategy only uses the loss of the vertices in the simplex, which in this case corresponds to the values $\ell(a, z), a \in \{1, ..., d\}$. Thus we have the following strategy, $\forall a \in \{1, ..., d\}$,

$$p_t(a) = \frac{\exp(-\eta \sum_{s=1}^{t-1} \ell(a, z_s))}{\sum_{i=1}^{d} \exp(-\eta \sum_{s=1}^{t-1} \ell(i, z_s))}$$

Then the following theorem directly follows from Theorem 7 and Lemma 3, sampling $a_t \sim p_t$.

Theorem 6. *For any loss with values in* $[0, 1]$*, the finite Exp strategy with parameter* $\eta = 2\sqrt{2\frac{\log d}{n}}$ *satisfies with probability at least* $1 - \delta$*:*

$$R_n \leq \sqrt{\frac{n \log d}{2}} + \sqrt{\frac{n \log \delta^{-1}}{2}} \tag{5}$$

Proof. In the Lemma 3 we note that the part on the L.H.S of the inequality is the cumulative regret (mentioned in the Section 1), that is,

$$R_n = \sum_{t=1}^{n} \ell(a_t, z_t) - \min_{a \in \mathcal{A}} \sum_{t=1}^{n} \ell(a, z_t)$$

and$\sqrt{\frac{n \log d}{2}}$, as part of the R.H.S of (5), is bounded by Theorem 1 since here we use $\eta = 2\sqrt{2\frac{\log d}{n}}$, that is,

$$\sum_{t=1}^{n} \bar{\ell}(p_t, z_t) - \min_{q \in \Delta_d} \sum_{t=1}^{n} \bar{\ell}(q, z_t) \leq \sqrt{\frac{n \log d}{2}}$$

Together we have

$$R_n \leq \sqrt{\frac{n \log d}{2}} + \sqrt{\frac{n \log \delta^{-1}}{2}}$$

□

3 Continuous exponential weights

Recall that in section 1.2, we showed that for problems involving the constantly rebalanced portfolio, we have the regret:

$$R_n = \sum_{s=1}^{n} -\log(a_s^T z_s) - \min_{a \in \mathcal{A}} \sum_{s=1}^{n} -\log(a^T z_s)$$

which leads to the log-loss $\ell(a, z) = -log(a^T z)$ as defined in Proposition 1. In fact, this regret can be viewed as the cumulative regret for the online optimization problem on a $(d-1)$-simplex with the log-loss.

However, as is also shown in Proposition 1, the log-loss $(a, z) \in \mathbb{R}_+^d \times \mathbb{R}_+^d \to -\log(a^T z)$ takes unbounded values and it has unbounded gradient, even when restricted to the $(d-1)$-simplex, which differs from linear loss. Note that the exponential strategy proposed previously works for the subdifferentiable losses with bounded subgradient, but for this example since the subgradients are unbounded, we would need a more general strategies that can work with infinite number of points. See [Bubeck Ch.3] [2] for details. Consider continuous Exp strategy defined as follows:

$$a_t = \int_{a \in \mathcal{A}} \frac{w_t(a)}{W_t} a da,$$

where

$$w_t(a) = \exp(-\eta \sum_{s=1}^{t-1} \ell(a, z_s)), \quad W_t = \int_{a \in \mathcal{A}} w_t(a) da,$$

and $\eta > 0$ a fixed parameter. So for each point $a \in \mathcal{A}$, a weight $w_t(a)$ is defined and corresponding weighted average a_t are computed.

The regret bound of continuous Exp with exp-concave losses is given:

Theorem 7. *For any σ-exp-concave loss ℓ, i.e, $(p,z) \in \mathbb{R}_+^d \times \mathbb{R}_+^d \mapsto \exp(-\sigma\ell(p,z))$ is concave, the Continuous Exp strategy has the following regret bound with parameter $\eta = \sigma$:*

$$R_n \leq \frac{1}{\sigma} d(1 + \log(n+1)).$$

The proof for this theorem can be found in the paper by Hazan et al. [4]. Note this is not the expert regret.

4 Online Gradient Descent

The Online Gradient Descent (OGD) is a strategy that can be applied to any closed convex set $\mathcal{A}$ and subdifferentiable loss ℓ. The strategy can be described as follows: Suppose a starting point $a_1 \in \mathcal{A}$, for $t \leq 1$,

$$w_{t+1} = a_t - \eta \nabla \ell(a_t, z_t), \qquad (4.1)$$
$$a_{t+1} = \arg\min_{a \in \mathcal{A}} ||w_{t+1} - a||_2 . \qquad (4.2)$$

Comparing with the Exponential strategy from the previous section, OGD requires stronger conditions, but is much simpler with better regret bounds as shown in Theorem 8 below. First we'll need the definition of subgradient and subdifferentiable:

Definition 3. *A vector $g \in \mathbb{R}^n$ is a subgradient of $f : \mathbb{R}^n \mapsto \mathbb{R}$ at $x \in \mathbb{R}^n$ if for all $z \in \mathbb{R}^n$, $f(z) \geq f(x) + g^T(z-x)$. If f is differentiable at x, then $g = \nabla f(x)$ uniquely.*

Definition 4. *f is subdifferentiable if $\forall x \in \mathcal{X}$, there exists a subgradient $g \in \mathbb{R}^d$ such that*
$f(x) - f(y) \leq g^T(x-y), \forall y \in \mathcal{X}$

Theorem 8. *(OGD) For any closed convex action set $\mathcal{A}$ such that $||a||_2 \leq R, \forall a \in \mathcal{A}$, for any subdifferentiable loss with bounded subgradient $||\nabla\ell(a,z)||_2 \leq G, \forall (a,z) \in \mathcal{A} \times \mathcal{Z}$, the OGD strategy with parameters $\eta = \frac{R}{G\sqrt{n}}$ and $a_1 = 0$ satisfies:*

$$R_n \leq RG\sqrt{n}$$

Proof. Let $g_t = \nabla\ell(a_t, z_t)$ for $a \in \mathcal{A}$. Thus by definition of regret, we can have $R_n = \sum_{t=1}^n (\ell(a_t, z_t) - \ell(amt_z) = \sum_{t=1}^n g_t^T(a_t - a)$. By (4.1) from above, we can have $\eta g_t = (a_t - w_{t+1})$. It follows:

$$\begin{aligned} 2\eta g_t^T(a_t - a) =& 2(a_t - w_{t+1})^T(a_t - a) \\ =& ||a - a_t||_2^2 + ||a_t - wa_{t+1}||_2^2 - ||a - w_{t+1}||_2^2 \end{aligned} \qquad (6)$$

where again by (4.1), $||a_t - w_{t+1}||_2^2 = \eta^2 ||g_t||_2^2$; and since $\mathcal{A}$ is a convex set, we have $||a - w_{t+1}||_2^2 \geq ||a - a_{t+1}||_2^2$ Therefore, by summing both sides of (6), we obtain:

$$\begin{aligned} 2\eta \sum_{t=1}^n g_t^T(a_t - a) \leq& ||a - a_1||_2^2 + \eta^2 \sum_{t=1}^n ||g_t||_2^2 \\ \leq& R^2 + \eta^2 G^2 n = 2R^2 \end{aligned}$$

$$\Rightarrow R_n \leq RG\sqrt{n}$$

□

For example, consider a classification problem using linear SVM with the following formula:

$$y = sign\{w_a^T x + b_a\}$$

where (w, b) are the parameters for the hyperplane. Hinge loss, or max-margin loss, is commonly used in SVM with loss function

$$\ell(a,(x,y)) = max(0, 1 - y(w_a^T x + b_a))$$

Let $w^* = (w_1, ..., w_t, b)$, we can say that the objective of a linear classifier is to find the hyperplane that minimizing the hinge loss, or

$$w^* = \min_{w,b} \sum_{t=1}^{n} max\{0, 1 - y(w_a^T x + b_a)\}$$

The regret is given by

$$R_n = \frac{1}{n} \sum_{t=1}^{n} max\{0, 1 - y(w_{a_t}^T x + b_{a_t})\} - \inf_{a \in \mathcal{A}} \sum_{t=1}^{n} max\{0, 1 - y(w_a^T x + b_a)\}$$

Let $y \in \{-1, 1\}$,with a feasible action set that $||a||_2 \leq R = G$, the subgradients of the hinge loss is given by:

$$\nabla \ell(a_t, z_t) = \partial_w \max\{0, 1 - y_n(wx_n + b)\} = \begin{cases} -y_n x_n, & \text{if } y_n(wx_n + b) \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

we have $||\nabla \ell(a_t, z_t)||_2 \leq ||a||_2 \leq R$. By applying the theorem above, we can find an upper bound for this regret

$$R_n \leq RG\sqrt{n} = R^2\sqrt{n}$$

References

[1] K.Azuma. *Weighted sums of certain dependent random variables.*. Tohoku Mathematical Journal, Volume 19, Number 3 (1967), 357-367.

[2] S. Bubeck. *Introduction to Online Optimization.*. Lecture Notes, 2011.

[3] A. Kalai and S. Vempala. *Universal portfolios.*. Math. Finance, 1:1-29, 1991

[4] E. Hazan, A. Agarwal, and S. Kale. *Logarithmic regret algorithms for online convex optimization.*. optimization. Machine Learning, 69:169-192, 2007.

[5] L. Lovasz and S. Vempala. *Fast algorithms for logconcave functions: sampling, rounding, integration and optimization.*. Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 57-68, 2006.