Sequential Learning with full information (Part 2)

Pierre Gaillard

1 Reminder from last week

We recall the setting of online convex optimization with full information in Figure 1.

At each time step $t = 1, \dots, T$
- the player observes a context $x_t \in \Theta$ (optional step)
- the player chooses an action $\theta_t \in \Theta$ (compact decision/parameter set);
- the environment chooses a loss function $\ell_t : \Theta \to [0, 1]$;
- the player suffers loss $\ell_t(\theta_t)$ and observes
  - the losses of every actions: $\ell_t(\theta)$ for all $\theta \in \Theta$ $\rightarrow$ full-information feedback
  - the loss of the chosen action only: $\ell_t(\theta_t)$ $\rightarrow$ bandit feedback.

The goal of the player is to minimize his cumulative loss:

$$\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t)\,.$$

Figure 1: Setting of an online learning problem/online convex optimization

Last week, we saw
- In the case of linear loss function defined by a loss vector $\ell_t(\theta) = \theta^\top g_t$ with $g_t \in [-1, 1]^K$, we defined the Exponentially Weighted Average (EWA) forecaster which updates

$$p_t(k) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(k)}}{\sum_{j=1}^{K} e^{-\eta \sum_{s=1}^{t-1} g_s(j)}}\,. \tag{EWA}$$

  It achieves a cumulative regret R_T of order $\sqrt{T \log K}$ when the set of actions is the K-dimensional simplex and when the loss functions are linear.
- In the setting of prediction with expert advice, the same upper-bound holds for EWA played with loss vectors $g_t = (\ell(x_t(1), y_t), \dots, \ell(x_t(K), y_t))$ for convex loss functions ℓ. For exp-concave loss functions we proved a constant regret.

In this class, we will generalize the results of last week to non-linear loss functions ℓ_t. We will consider first the simplest case of finite $\Theta = \{1, \dots, K\}$ with arbitrary bounded loss functions, then the case of the simplex $\Theta = \Delta_K$ with convex loss functions ℓ_t and finally any compact convex set Θ.

2 Random EWA

Setting: Θ finite, non-convex loss functions $\ell_t : \Theta \to [-1, 1]$.

In this section, we consider a *finite set of decision* $\Theta = \{1, \dots, K\}$ and we assume that the player is restricted to play an action in Θ. In other words, the player cannot play convex combinations of the actions as it was done for prediction with expert advice. For instance, we may want to build a recommender system to recommend movies to customers. The loss function are *arbitrary bounded loss functions* $\ell_t : \Theta \to [-1, 1]$.

Need of a random strategy The following proposition shows that the choice θ_t cannot be deterministic in this setting. Otherwise, the adversary may fool the player by taking ℓ_t depending on θ_t.

Proposition 1. *Any deterministic algorithm may incur a linear regret. In other words, we can find some sequence of losses ℓ_t such that $R_T \gtrsim T$.*

Proof. Since θ_t is deterministic, the loss function ℓ_t can depend on θ_t. We then choose $\ell_t(\theta_t) = 1$ and $\ell_t(\theta) = 0$ for $\theta \neq \theta_t$. Then one of the chosen actions was picked less then T/K times so that $\max_{1 \le k \le K} \ell_t(k) \le T/K$. Therefore, $R_T \ge (1 - 1/K)T$. □

From the above proposition, we see that the strategy of the learner needs to be random. Therefore, instead of choosing an action in $\{1, \dots, K\}$, the player chooses a probability distribution $p_t \in \Delta_K := \{p \in [0,1]^K : \sum_k p_k = 1\}$ and draws $\theta_t \sim p_t$. And we recover the setting with actions played in the simplex Δ_K.

A random regret The regret R_T will be here a random quantity that depends on the randomness of the algorithm (and eventually of the data). We will thus focus on upper-bounding the regret:

- with high-probability: $R_T \le \varepsilon$ with probability at least $1 - \delta$;
- in expectation: $\mathbb{E}[R_T] \le \varepsilon$.

From high-probability bound to expected bound. Note that since the losses are bounded in $[0, 1]$ a bound in high probability entails a bound in expectation. If $R_T \le \varepsilon$ with probability at least $1 - \delta$, then

$$\mathbb{E}[R_T] \le \mathbb{E}\big[R_T | R_T \le \varepsilon\big]\mathbb{P}(R_T \le \varepsilon) + \mathbb{E}\big[R_T | R_T \ge \varepsilon\big]\mathbb{P}(R_T \ge \varepsilon) \le \varepsilon + T\delta\,. \tag{1}$$

Another useful (and often better) tool to transform a high-probability bound into a bound in expectation is the inequality $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \ge \varepsilon) d\varepsilon$ for nonnegative random variable X.

From expected bound to high-probability bound. On the other hand, since the losses are bounded, using Hoeffding's inequality a bound in expectation entails a bound in high probability at the cost of an additive term of order $\sqrt{T \log(1/\delta)}$ in the regret.

Proposition 2. *Drawing $\theta_t \sim p_t$ where p_t are chosen by EWA satisfies the expected regret*

$$\mathbb{E}\big[R_T\big] = \mathbb{E}\left[\sum_{t=1}^T \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^T \ell_t(\theta)\right] \le 2\sqrt{T \log K}$$

for η well tuned.

Exercise 2.1. Using Hoeffding's inequality, provide a bound on the regret R_T with probability $1 - \delta$.

Proof. Using $g_t = \big(\ell_t(1), \dots, \ell_t(K)\big) \in [-1, 1]^K$, from Theorem 1 of last class, we have

$$\sum_{t=1}^T p_t \cdot g_t - \min_{\theta \in \Theta} \sum_{t=1}^T \ell_t(\theta) \le 2\sqrt{T \log K}\,.$$

It suffices then to take the expectation and remark that

$$\mathbb{E}\big[\ell_t(\theta_t)\big] = \mathbb{E}\big[\mathbb{E}[\ell_t(\theta_t) | p_t]\big] = \mathbb{E}\big[p_t \cdot g_t\big]\,.$$

□

It is worth pointing out that we did not make any assumption on the loss function ℓ_t beside boundedness. In particular, it can be non-convex.

Example 2.1 (Online classification)**.** Assume that you may want to predict a sequence of labels $y_1, \ldots, y_T \in \{0,1\}$ (such as spams) based on expert advice $x_t(k) \in \{0,1\}$ (such as different spam detectors). Then, using the losses $\ell_t(k) = \mathbb{1}_{x_t(k)\neq y_t}$, EWA ensures

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{\theta_t \neq y_t} - \min_{1\leq k \leq K} \sum_{t=1}^{T} \mathbb{1}_{x_t(k)\neq y_t}\right] \leq 2\sqrt{T \log K}\,.$$

Hence, the expected number of mistakes of the algorithms will not be much larger than the one of the best expert. This is valid though the loss function is nonconvex.

3 The Gradient Trick (from linear to convex losses)

Setting: simplex decision set $\Theta = \Delta_K$, convex and differentiable loss functions

In this section, we consider the simplex decision set $\Theta = \Delta_K$ and thus we will denote by p_t (instead of θ_t) the actions played by the player. Moreover, we assume the losses to be convex and Lipschitz

$$\forall p_t \in \Theta, \qquad \big\|\nabla \ell_t(p_t)\big\|_\infty \leq G.$$

We will see a simple trick, so-called *the gradient trick* that allows to extend the results we saw for linear losses to convex losses. The resulted algorithm is called the Exponentiated Gradient forecaster (EG). It consists in playing EWA with the gradients $g_t = \nabla \ell_t \in [-G, G]^K$ as loss vectors.

Theorem 1. *Let $T \geq 1$. For all sequences of convex differentiable losses $\ell_1, \ldots, \ell_T : \Theta \to \mathbb{R}$ with bounded gradient $\max_{p\in\Theta} \|\nabla \ell_t(p)\|_\infty \leq G$, EWA applied with $g_t = \nabla \ell_t$ achieves the regret bound*

$$R_T \stackrel{def}{=} \sum_{t=1}^{T} \ell_t(p_t) - \min_{p\in\Theta} \sum_{t=1}^{T} \ell_t(p) \leq \eta G^2 T + \frac{\log K}{\eta}\,. \tag{2}$$

Therefore, for the choice $\eta = \frac{1}{G}\sqrt{\frac{\log K}{T}}$, EWA satisfies the regret bound $R_T \leq 2G\sqrt{T \log K}$.

Proof. Therefore, applying the regret bound of EWA (see Theorem 1 of last class) we get

$$\sum_{t=1}^{T} p_t \cdot g_t - \min_{p\in\Delta_K} \sum_{t=1}^{T} p \cdot g_t \leq \eta \sum_{t=1}^{T}\sum_{k=1}^{K} p_t(k) g_t(k)^2 + \frac{\log K}{\eta}\,.$$

Remark that the theorem also holds for loss vectors $g_t \in [-G, G]^K$ as soon as $\eta \leq 1/G$. Upper-bounding $g_t(j)^2 \leq \|\nabla \ell_t(p_t)\|_\infty^2 \leq G^2$, substituting $g_t = \nabla \ell_t(p_t)$, this yields for all $p \in \Delta_K$

$$\sum_{t=1}^{T} p_t \cdot \nabla \ell_t(p_t) - p \cdot \nabla \ell_t(p_t) \leq \eta T G^2 + \frac{\log K}{\eta}\,.$$

But by convexity of the losses, we have the gradient inequality

$$\ell_t(p_t) - \ell_t(p) \leq (p_t - p) \cdot \nabla \ell_t(p_t)\,,$$

which yields

$$\sum_{t=1}^{T} \ell_t(p_t) - \ell_t(p) \leq \eta T G^2 + \frac{\log K}{\eta}\,.$$

The proof is concluded by optimizing $\eta = \frac{1}{G}\sqrt{\frac{\log K}{T}}$. ∎

Example 3.1 (Prediction with expert advice (continued))**.** In prediction with expert advice, a sequence of observations $y_1, \dots, y_T \in [0,1]$ is to be predicted with the help of K expert advice $x_t(k) \in [0,1]$ for $1 \le k \le K$. The learner predict $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$ and suffers a loss $\ell(\widehat{y}_t, y_t)$. If the loss function is convex and Lipschitz in its first argument we can apply Theorem 1 with $\ell_t : p \mapsto \ell(p \cdot x_t, y_t)$. For instance, with the absolute loss, $G = 1$ and EG satisfies a bounded regret with respect to any fixed convex combination of experts:

$$\sum_{t=1}^T |\widehat{y}_t - y_t| - \min_{p \in \Theta} \sum_{t=1}^T \Big| p \cdot x_t - y_t \Big| \le 2\sqrt{T \log K} \,.$$

Hence, on the long run we perform as good as the best convex combination of the experts which may outperform the best expert. This may leads to much better performance than a simple EWA on the experts if

$$\min_{p \in \Theta} \sum_{t=1}^T \Big| p \cdot x_t - y_t \Big| \ll \min_{k \in [K]} \sum_{t=1}^T \Big| x_t(k) - y_t \Big|.$$

Convex hull of finite point set It is worth pointing out that the simplex decision set Δ_K can be generalized with any convex hull of a finite point set $S = \{\theta(1), \dots, \theta(K)\}$:

$$\mathrm{Conv}(S) = \left\{ \sum_{i=1}^K p_i \theta(i) : \forall i, p_i > 0 \text{ and } \sum_{i=1}^K p_i = 1 \right\} .$$

Transforming the loss functions, EWA can be applied to compete with such sets as shown by the theorem bellow.

Theorem 2. *Let $T \ge 1$. Let $\Theta \subset \mathbb{R}^d$ be a convex set and $S = \{\theta(1), \dots, \theta(K)\} \in \Theta^K$ with diameter $D \ge \max_{i,j} \|\theta(i) - \theta(j)\|_1$. Let $\ell_1, \dots, \ell_T : \Theta \to \mathbb{R}$ be an arbitrary sequence of convex differentiable losses with bounded gradient $\max_{\theta \in \Theta} \|\nabla \ell_t(\theta)\|_\infty \le G$. Then, EWA applied with $g_t = \nabla \tilde{\ell}_t$ where $\tilde{\ell}_t : p \mapsto \ell_t \left(\sum_{i=1}^K p(i)\theta(i) \right)$ achieves the regret bound*

$$R_T \stackrel{def}{=} \sum_{t=1}^T \ell_t(\theta_t) - \min_{\theta \in Conv(S)} \sum_{t=1}^T \ell_t(\theta) \le 2GD\sqrt{T \log K} \,,$$

where $\theta_t = \sum_{k=1}^K p_t(k)\theta(k)$

Such a trick can be used for instance to compete with the ℓ_1-balls using $S = \{\theta \in \mathbb{R}^d : \|\theta\|_1 = R, \|x\|_0 = 1\}$. Since ℓ_p-balls are contained into the ℓ_1-ball (of possibly larger radius depending on p) this can also be used to compete against any ℓ_p-ball for $p \ge 1$. This trick was introduced by Kivinen and Warmuth [1997] for the EG$\pm$ forecaster.

4 Discretized EWA

Setting: general compact decision set, β-Hölder loss functions

In this section, we aim at designing a procedure for general compact decision set Θ. We will assume for simplicity that $\Theta \subset \mathbb{R}^d$ with $\max_{\theta, \theta' \in \Theta} \|\theta - \theta'\| \le D$, where $\|\cdot\|$ denotes the Euclidean norm. If the loss functions ℓ_t are β-Hölder, i.e.,

$$\Big| \ell_t(\theta) - \ell_t(\theta') \Big| \le c \|\theta - \theta'\|^\beta$$

there exists a simple solution: approximate Θ with a finite discretization grid Θ_ε and apply EWA on Θ_ε. If Θ or the losses are non-convex, one needs to use the random EWA (see Section 2) and bound the regret with high-probability. For convenience, we will assume Θ and the loss functions ℓ_t to be convex so that the algorithm can play convex combinations of points in Θ_ε and all quantities are deterministic.

Lemma 1. *Let $\varepsilon > 0$. Let $\Theta \subset \mathbb{R}^d$ such that $\max_{\theta,\theta' \in \Theta} \|\theta - \theta'\| \leq D$. Then, there exists $\Theta_\varepsilon \subset \Theta$ such that*

$$\mathrm{Card}(\Theta_\varepsilon) \lesssim \left(\frac{D}{\varepsilon}\right)^d \qquad \textit{and} \qquad \forall x \in \Theta, \exists x' \in \Theta_\varepsilon \quad \|\theta - \theta'\| \leq \varepsilon\,,$$

where $\lesssim$ denotes a rough inequality (up to multiplicative constants and logarithmic terms).

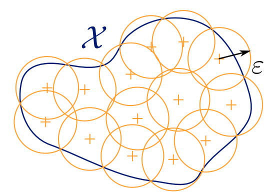
Remark. Remark that a set finite Θ_ε which approximate Θ at radius ε, is called an ε-covering of Θ. The cardinal of the smallest ε-covering is called the *covering number* of Θ. This cardinal is heavily used in theory to analyze the complexity of general spaces Θ. It heavily differentiates parametric spaces with covering number of order $(1/\varepsilon)^d$ with nonparametric spaces (spaces of functions) for which the logarithm of the covering number (or metric entropy) is of order $(1/\varepsilon)^d$.

Proof sketch. We only provide the high-level idea of the proof. First, by properties of the Lebesgue measure in d-dimension, denoting $\mathcal{B}_2(r)$ is the ℓ_2-ball of radius $r > 0$, we have

$$\mathrm{Vol}\big(\mathcal{B}_2(r)\big) = \frac{\pi^{d/2}}{\Gamma(n/2+1)} r^d\,,$$

where Γ is the Euler's gamma function. Therefore,

$$\mathrm{Vol}(\Theta) \leq \mathrm{Vol}\big(\mathcal{B}_2(D/2)\big) = \left(\frac{D}{2\varepsilon}\right)^d \mathrm{Vol}\big(\mathcal{B}_2(\varepsilon)\big)\,,$$



and thus approximatively $\left(\frac{D}{2\varepsilon}\right)^d$ balls of radius ε are sufficient to cover Θ.

□

Theorem 3 (Discretized EWA)**.** *Let $T \geq 1$, $\varepsilon, D > 0$. Let Θ be a compact convex subset of $\mathbb{R}^d$ such that $\max_{\theta,\theta' \in \Theta} \|\theta - \theta'\| \leq D$. Let Θ_ε be an ε-covering of Θ with smallest cardinal. Then, for all sequences of β-Hölder convex losses $\ell_1, \ldots, \ell_T : \Theta \to [0,1]$, EWA played on the finite set of action Θ_ε with optimized η satisfies the regret bound*

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^T \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^T \ell_t(\theta) \lesssim \sqrt{Td\Big(\log D + \frac{1}{\beta}\log(cT)\Big)}\,.$$

Exercise 4.1. Provide a bound on the expected regret for random EWA when the losses and the decision set are non-convex.

Proof. Let $K = \mathrm{Card}(\Theta_\varepsilon)$. Let us order the elements of $\Theta_\varepsilon = \{\theta(1), \ldots, \theta(K)\}$. Therefore, at time $t \geq 1$, EWA chooses a weight vector $p_t \in \Delta_K$ and predict the weighted average $\theta_t = \sum_{k=1}^K p_t(k)\theta(k) \in \Theta$. Applying the regret bound of EWA, we get

$$\sum_{t=1}^T \sum_{k=1}^K p_t(k)\ell_t(\theta(k)) - \min_{1 \leq j \leq K} \sum_{t=1}^T \ell_t(\theta(j)) \leq 2\sqrt{T \log K}\,. \tag{3}$$

Let $\theta^* \in \Theta$ and $\theta(k^*) \in \Theta_\varepsilon$ such that $\|\theta^* - \theta(k^*)\| \le \varepsilon$. Because the losses are β-Hölder and convex, we have

$$
\begin{aligned}
R_T \quad & = \quad \sum_{t=1}^T \ell_t(\theta_t) - \ell_t(\theta^*) \\
& \overset{\text{Convexity}}{\le} \quad \sum_{t=1}^T \sum_{k=1}^K p_t(k)\ell_t(\theta(k)) - \ell_t(\theta^*) \qquad \leftarrow \theta_t = \sum_{k=1}^K p_t(k)\theta(k) \\
& \le \quad \sum_{t=1}^T \sum_{k=1}^K p_t(k)\ell_t(\theta(k^*)) - \ell_t(\theta(k^*)) + \sum_{t=1}^T \big|\ell_t(\theta(k^*)) - \ell_t(\theta^*)\big| \\
& \overset{(3)}{\le} \quad 2\sqrt{T\log K} \quad + \quad T \max_{1\le t\le T} \big|\ell_t(\theta^*) - \ell_t(\theta(k^*))\big| \\
& \overset{\beta\text{-Hölder}}{\le} \quad 2\sqrt{T\log K} + cT\varepsilon^\beta \\
& \overset{\text{Lem. 1}}{\lesssim} \quad \sqrt{Td\log\Big(\frac{D}{\varepsilon}\Big)} + cT\varepsilon^\beta\,.
\end{aligned}
$$

Optimizing $\varepsilon^\beta = 1/cT$, hence $\varepsilon = (cT)^{-1/\beta}$, we get

$$
R_T \lesssim \sqrt{Td\Big(\log D + \frac{1}{\beta}\log(cT)\Big)}\,.
$$

□

Though this algorithm is theoretically convenient since it can deals with general compact sets Θ and general loss functions (which can be non-convex and non-differentiable). It suffers two considerable drawbacks:

- *computational complexity*: the algorithm needs to consider a discretization space of cardinal $(X/\varepsilon)^d$ which is of order $O(T^{d/\beta})$. This is prohibitive in practice.
- *bad regret dependence on the dimension*: the regret bound is of order $O(\sqrt{dT\log T})$. We will see how to have no dependence on d when Θ is bounded in ℓ_2-norm.

5 Online gradient descent

Setting: convex differentiable Lipschitz loss function, convex and compact decision set Θ

In this section, we provide another algorithm that solves the drawbacks of the discretized EWA seen in the previous section. This algorithm is called, Online Gradient Descent, and is due to Zinkevich [2003]. It is an online variant of the well-known Gradient Descent algorithm in optimization.

Online Gradient Descent (OGD)

Parameter: $\eta > 0$

Initialize: $\theta_1 \in \Theta$ arbitrarily chosen

For $t = 1, \dots, T$

- select θ_t; incur loss $\ell_t(\theta_t)$ and observe $\ell_t : \Theta \to [0,1]$;
- compute the gradient $\nabla \ell_t(\theta_t)$
- update

$$
\theta_{t+1} = \Pi_\Theta\Big(\theta_t - \eta\nabla\ell_t(\theta_t)\Big)\,.
$$

where Π_Θ is the Euclidean projection onto Θ.

Theorem 4. *Let $D, G, \eta > 0$. Assume that $\max_{\theta,\theta'\in\Theta} \|\theta - \theta'\| \leq D$ and. Then for any sequence $\ell_1, \dots, \ell_T$ of convex differentiable loss functions such that $\max_{\theta\in\Theta} \|\nabla \ell_t(\theta)\| \leq G$, the regret of OGD satisfies*

$$\sum_{t=1}^T \ell_t(\theta_t) - \min_{\theta\in\Theta} \sum_{t=1}^T \ell_t(\theta) \leq \frac{D^2}{2\eta} + \frac{\eta}{2} G^2 T\,.$$

In particular, for $\eta = \frac{D}{G\sqrt{T}}$, we have $R_T \leq DG\sqrt{T}$.

Exercise 5.1. Prove an upper-bound on the regret of OGD
  a) when η is calibrated with a doubling trick.
  b) when η is calibrated using a time-varying parameter η_t

Exercise 5.2. Prove an upper-bound on the regret of OGD with respect to any sequence of points $\theta_1^*, \dots, \theta_t^* \in \Theta$ such that $\sum_{t=2}^T \|\theta_t^* - \theta_{t-1}^*\| \leq X$

$$\sum_{t=1}^T \ell_t(\theta_t) - \sum_{t=1}^T \ell_t(\theta_t^*) \leq \quad \dots$$

Remark. Assume that $\Theta = \Delta_K$ is the simplex and the loss functions are sub-differentiable convex functions with $\|\nabla \ell_t\|_\infty \leq G_\infty$. Then both EG and OGD are possible algorithms (see Theorems 1 and 4). We saw in Theorem 1 that EG has a regret bound $R_T \leq 2G_\infty\sqrt{T\log K}$. In this case, for all $p, p' \in \Delta_K$

$$\|p - p'\| = \sum_{k=1}^K \big(p(i) - p'(i)\big)^2 \leq \sum_{i=1}^K \big|p(i) - p'(i)\big| \leq \sum_{i=1}^K p(i) + p'(i) = 2\,,$$

and $\|\nabla \ell_t(p)\| \leq \sqrt{K}\|\nabla \ell_t(p)\|_\infty \leq \sqrt{K}G_\infty$. Therefore, the regret of OGD is upper-bounded by $R_t \leq G_\infty\sqrt{2KT}$. To summarize

$$\text{EG:}\quad R_T \leq 2G_\infty\sqrt{T\log K} \qquad \text{and} \qquad \text{OGD:}\quad R_T \leq \sqrt{2KT}\,.$$

The dependence on K of OGD is suboptimal in this case. This is solved by OMD, a generalization of both algorithms.

Proof of Theorem 4. Let $\theta^* \in \arg\min_{\theta\in\Theta} \sum_{t=1}^T \ell_t(\theta)$ and denote $z_{t+1} = \theta_t - \eta\nabla\ell_t(\theta_t)$ so that by definition of θ_{t+1} in the algorithm, we have $\theta_{t+1} = \Pi_\Theta(z_{t+1})$. By convexity, the regret can be upper-bounded as

$$\begin{aligned} R_T \stackrel{\text{def}}{=} \sum_{t=1}^T \ell_t(\theta_t) - \ell_t(\theta^*) \quad &\leq \quad \sum_{t=1}^T \nabla\ell_t(\theta_t)\cdot(\theta_t - \theta^*) \\ &= \quad \frac{1}{\eta}\sum_{t=1}^T (z_{t+1} - \theta_t)\cdot(\theta_t - \theta^*) \qquad \leftarrow \nabla\ell_t(\theta_t) = \frac{z_{t+1} - \theta_t}{\eta}\,. \end{aligned}$$

Then, we use the equality $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2x\cdot y$ for all $x, y \in \Theta$ so that

$$x\cdot y = \frac{\|x\|^2 + \|y\|^2 - \|x-y\|^2}{2}\,.$$

Applying it with $x = z_{t+1} - \theta_t$ and $y = \theta_t - \theta^*$ en substituting into the above regret bound, this yields

$$R_T \leq \frac{1}{2\eta}\sum_{t=1}^T \big(\|z_{t+1} - \theta_t\|^2 + \|\theta^* - \theta_t\|^2 - \|z_{t+1} - \theta^*\|^2\big)$$

Then, using $\|z_{t+1} - \theta_t\| = \eta\|\nabla\ell_t(\theta_t)\| \leq \eta G$ and $\|\theta^* - \theta_t\| \leq \|\theta^* - z_t\|$ because Θ is convex and $\theta_t = \Pi_\Theta(z_t)$, we get

$$R_T \leq \frac{1}{2\eta}\sum_{t=1}^{T}\left(\eta^2 G^2 + \|\theta^* - z_t\|^2 - \|z_{t+1} - \theta^*\|^2\right).$$

The last terms telescope, therefore summing over t concludes the proof

$$R_T \leq \frac{\eta G^2 T}{2} + \frac{\|\theta^* - \theta_1\|^2}{2\eta} \leq \frac{\eta G^2 T}{2} + \frac{D^2}{2\eta}.$$

□

Online Mirrored Descent

Online Mirrored Descent (OMD) is a generalization of OGD to better exploit the geometry of the decision space Θ. OMD is the online counterpart of the *Mirrored Descent* algorithm from convex optimization. The generality of OMD comes from the updates being performed into a dual space which is defined by a convex differentiable regularization function $R : \Theta \to \mathbb{R}$.

Before stating the algorithm, we need to define the Bregman divergence.

Definition 1 (Bregman divergence)**.** *For any continuously differentiable convex function R, the Bregman divergence with respect to R is defined as*

$$D_R(x||y) \leq R(x) - R(y) - \nabla R(y)\cdot(x-y) \quad \forall x, y \in \Theta.$$

It is the difference between the value of the regularization function at x and the value of its first order Taylor approximation. It is nonnegative but not symmetric. Online Mirrored Descent is then defined as follows.

Online Mirrored Descent (OMD)

Parameters: $\eta > 0$, regularization function R
Initialize: $z_1 \in \mathbb{R}^d$ such that $\nabla R(z_1) = 0$ and $\theta_1 = \arg\min_{\theta\in\Theta} B_R(\theta||y_1)$

For $t = 1, \dots, T$
- select θ_t; incur loss $\ell_t(\theta_t)$ and observe $\ell_t : \Theta \to [0,1]$;
- compute the gradient $\nabla\ell_t(\theta_t)$
- update z_t such that

$$\nabla R(z_{t+1}) = \nabla R(\theta_t) - \eta\nabla\ell_t(\theta_t).$$

- project according to the Bregman divergence

$$\theta_{t+1} \in \operatorname*{arg\,min}_{\theta\in\Theta} D_R(\theta||z_{t+1}).$$

Theorem 5. *Let $t \geq 1$. Let Θ be a compact and convex set. Then, for all sequences of convex subdifferentiable loss functions $\ell_1, \dots, \ell_T : \Theta \to [0,1]$, the regret of OMD is upper-bounded as*

$$\sum_{t=1}^{T}\ell_t(\theta_t) - \min_{\theta\in\Theta}\sum_{t=1}^{T}\ell_t(\theta) \leq \frac{D}{\eta} + \frac{1}{\eta}\sum_{t=1}^{T} D_{R^*}\big(\nabla R(\theta_t) - \eta\nabla\ell_t(\theta_t)||\nabla R(\theta_t)\big)$$

where $D \geq \max_{\theta\in\Theta}|R(\theta)|$ and R^* is the Fenchel conjugate of R defined as $R^*(z) \overset{def}{=} \max_{\theta\in\Theta}\big\{\theta\cdot z - R(\theta)\big\}$.

The proof can be found for instance in Bubeck et al. [2012]. EG and OGD are two particular cases of Online Mirror Descent.

Example 5.1 (Balls in $\mathbb{R}^d$ = OGD)**.** If $\Theta \subset \mathbb{R}^d$, we can choose $R(x) = \frac{1}{2}\|x\|^2$. Then $\nabla R(x) = x$ and $D_R(x||y) = \frac{1}{2}\|x-y\|^2$. Therefore, the update of OMD becomes $y_{t+1} = \theta_t - \eta \nabla \ell_t(\theta_t)$ and $\theta_{t+1} = \Pi_\Theta(y_{t+1})$. We recover the online gradient descent algorithm.

Example 5.2 (Simplex = EG)**.** If $\Theta = \Delta_K$. We can choose the negative entropy

$$R(x) = \sum_{i=1}^{K} x(i) \log x(i)\,.$$

In this case, $\nabla R(x)_i = 1 + \log x(i)$ and the Bregman Divergence is $D_R(x||y) = \sum_{i=1}^K x(i)\log(x(i)/y(i))$ also known as the Kullback-Leibler divergence. The update of OMD is then

$$1 + \log(y_{t+1}(i)) = 1 + \log \theta_t(i) - \eta g_t(i)\,,$$

where $g_t = \nabla \ell_t(\theta_t) \in \mathbb{R}^K$. This can be rewritten

$$y_{t+1}(i) = \theta_t(i) e^{-\eta[\nabla \ell_t(\theta_t)]_i}.$$

The projection to the simplex is a simple renormalization (left as exercise), we thus get

$$\theta_{t+1}(i) = \frac{\theta_t(i) e^{-\eta g_t(i)}}{\sum_{k=1}^K \theta_t(k) e^{-\eta g_t(k)}}\,,$$

and we recover the update of EG (i.e., EWA applied with the gradient trick $g_t = \nabla \ell_t(\theta_t)$).

References

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Jyrki Kivinen and Manfred K Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *information and computation*, 132(1):1–63, 1997.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936, 2003.