Online learning on a continuum

Walid Krichene

December 5, 2014

Abstract

We study a sequential decision problem on a subset $S \subset \mathbb{R}^n$. A decision maker chooses, on iteration $t$, a probability distribution $\pi(t)$ over $S$, then discovers a bounded loss function $\ell(t) : S \rightarrow [0,M]$, and incurs the expectation $\mathbb{E}_{s \sim \pi(t)} \ell(t)(s)$. The cumulative regret of the decision maker is then $\sum_{t=1}^T \mathbb{E}_{s \sim \pi(t)}[\ell(t)(x)] - \inf_{s \in S} \sum_{\tau=1}^T \ell(\tau)(s)$. We investigate conditions under which one can guarantee a sublinear regret. Previous studies consider the case where $S$ is convex: if the losses are convex, then a simple gradient descent algorithm guarantees a $O(\sqrt{t})$ regret, and if the losses are exp-concave, a generalized Hedge algorithm guarantees a $O(\log t)$ regret.

Building on this previous work, we relax the convexity assumption on $S$, and propose a generalized Hedge algorithm with a $O(\sqrt{t \log t})$ bound on the regret when the losses are Lipschitz (uniformly in time) and $S$ is uniformly fat (a weaker condition than convexity).

We compare our method to working with a finite cover of the set. In particular, we show that both guarantee a $O(\sqrt{t \log t})$ bound on the regret, but our method does not need to explicitly compute a cover.
1 Introduction

We consider the following online optimization problem: a decision maker has an action set $S \subset \mathbb{R}^n$, and chooses at each iteration $t$, a distribution $\pi(t)$ over $S$, then discovers a loss function $\ell(t) : S \rightarrow [0, M]$. The expected loss on iteration $t$ is then $E_{s \sim \pi(t)}[\ell(t)(s)]$. When choosing distribution $\pi(t)$, the decision maker only has access to the history of loss functions up to time $t - 1$. Thus an online learning algorithm (or strategy) is a function that maps a history of losses to a distribution on $S$ 

$$ \left( \ell(\tau) \right)_{1 \leq \tau \leq t - 1} \mapsto \pi(t) $$

When designing online learning algorithms, one usually seeks a guarantee on the cumulative regret, defined as $R(t) = E_{s \sim \pi(t)}[\ell(t)(s)] - \inf_{s \in S} \ell(t)(s)$. In particular, if the algorithm guarantees that for any sequence of loss functions, $\limsup \frac{R(t)}{t} \leq 0$, the algorithm is said to achieve sublinear regret.

The regret is a natural measure of performance in many online decision settings. In the context of repeated games, the regret measures how far the joint strategy of players is from the equilibrium set, see e.g. Hannan [1957]. Blackwell [1956] generalized the regret analysis to vector payoffs. In the context of convex online optimization, for example Zinkevich [2003], the instantaneous regret is closely related to the optimality conditions of the convex optimization problem. Regret is also a common measure of performance in stochastic bandit problems Lai and Robbins [1985], and adversarial online decision problems Cesa-Bianchi and Lugosi [2006].

The case where $S$ is finite is studied under different sets of assumptions, and many algorithms are known to have sublinear regret guarantees, such as the Hedge algorithm, perhaps one of the most popular algorithms, a.k.a. exponential weights algorithm Arora et al. [2012], experts algorithm or weighted majority algorithm Littlestone and Warmuth [1989] exponentiated gradient descent Kivinen and Warmuth [1997], entropic gradient descent Beck and Teboulle [2003] and many others Freund and Schapire [1999], Cesa-Bianchi and Lugosi [2006].
Assumptions | $\ell(t)$ convex | $\ell(t)$ $\alpha$-exp-concave | L Lipschitz on $v$ uniformly fat $S$
---|---|---|---
Method | gradient descent | Hedge | Hedge
$\eta_t$ | $\frac{1}{\sqrt{t}}$ | $\alpha$ | $\frac{1}{\sqrt{t}}$
$\frac{R(t)}{t}$ | $\mathcal{O}\left(\frac{1}{\sqrt{t}}\right)$ | $\mathcal{O}\left(\frac{\log t}{t}\right)$ | $\mathcal{O}\left(\sqrt{\frac{\log t}{t}}\right)$

Table 1: Summary of known upper bounds for different classes of losses.

An overview of basic results is given in Bubeck and Cesa-Bianchi [2012]. We focus our attention on the case where $S$ is infinite, particularly when it is a compact subset of $\mathbb{R}^n$. When $S$ is convex, Zinkevich [2003] shows that a gradient descent algorithm guarantees a $\mathcal{O}(\sqrt{t})$ upper bound on the regret when the loss functions are convex. Hazan et al. [2007] show that if the loss functions are $\alpha$ exp-concave, then a generalized Hedge algorithm with learning rate $\alpha$ achieves a $\mathcal{O}(\log t)$ regret.

We seek to relax the convexity assumptions on the action set and the loss functions, and consider a generalized Hedge algorithm with a decreasing sequence of learning rates. We start by proving a general bound in Section 2, which holds for any subset $S$ and any sequence of bounded loss functions. This basic bound does not guarantee, however, a sublinear bound on the cumulative regret, and one needs to make additional assumptions on $S$ or the loss functions. In Section 3, we specialize the bound to exp-concave losses and Lipschitz losses. We recover, as a special case, the result of Hazan et al. [2007], when $S$ is convex. Then we propose a sufficient condition on $S$ which guarantees a $\mathcal{O}(\sqrt{t \log t})$ regret for Lipschitz losses.

In Section 4, we compare this generalized Hedge algorithm to methods which compute a finite cover of the action set and applies an online learning algorithm on discrete sets. We argue that both achieve the same asymptotic upper bounds on the regret, and that the major difference is that one does not need to explicitly compute a cover to apply our proposed algorithm, even though the analysis is similar. The results are summarized in Tables 1 and 2.

### 2 A regret bound on metric spaces

Consider a compact subset $S$ of a metric space, and fix a $\sigma$-finite reference probability measure $\nu$ on $S$.

Let $\ell(t) : S \to [0, M]$ denote the loss function at iteration $t$, supposed to be continuous, and bounded by $M$, uniformly in $t$. We denote by $\Pi(t) \in \Delta_\nu(S)$ the decision maker’s distribution over
$S$ at iteration $t$, and by $\pi^{(t)}$ the density of $\Pi^{(t)}$ w.r.t. $\nu$.

The Hedge algorithm with initial density $\pi^{(0)} \in \Delta(S)$ and learning rates $(\eta_t)_t$ is defined in terms of the density functions as follows:

$$\pi^{(t+1)}(s) = \frac{1}{Z^{(t)}} \pi^{(0)}(s) \exp \left( -\eta_{t+1} \sum_{\tau=1}^{t} \ell^{(\tau)}(s) \right)$$

(1)

where $(\eta_t)$ is a decreasing sequence of learning rates, and $Z^{(t)}$ is the appropriate normalization constant, i.e., $Z^{(t)} = E_{s \sim \pi^{(0)}} \left[ \exp \left( -\eta_{t+1} \sum_{\tau=1}^{t} \ell^{(\tau)} \right) \right]$. The Hedge algorithm is summarized in Algorithm 1.

**Algorithm 1** Hedge algorithm with initial density $\pi^{(0)}$ and learning rates $(\eta_t)$.

```plaintext
for $t \in \mathbb{N}$ do
    Perform action $s \sim \pi^{(t)}$
    Observe loss function $\ell^{(t)}$
    Update $\pi^{(t+1)}(s) \propto \pi^{(0)}(s) \exp \left( -\eta_{t+1} \sum_{\tau=1}^{t} \ell^{(\tau)}(s) \right)$
end for
```

The instantaneous pseudo-regret function at iteration $t$ is

$$r^{(t)}(s) = E_{u \sim \pi^{(t)}}[\ell^{(t)}(u)] - \ell^{(t)}(s)$$

and the cumulative pseudo-regret is defined as

$$R^{(t)} = \sup_{s \in S} \sum_{\tau=1}^{t} r^{(\tau)}(s) = \sum_{\tau=1}^{t} E_{u \sim \pi^{(\tau)}}[\ell^{(\tau)}(u)] - \inf_{s \in S} \sum_{\tau=1}^{t} \ell^{(\tau)}(s)$$

The first term in the above expression is the expected cumulative loss of the decision maker, and will be denoted

$$L^{(t)} = \sum_{\tau=1}^{t} E_{u \sim \pi^{(\tau)}}[\ell^{(\tau)}]$$

The second term is the infimum of the cumulative loss function, which will be denoted

$$\mathcal{L}^{(t)}(s) = \sum_{\tau=1}^{t} \ell^{(\tau)}(s)$$

Since $S$ is compact and $\mathcal{L}^{(t)}$ is continuous the infimum of the cumulative loss function $\mathcal{L}^{(t)}$ is attained on $S$. Let

$$s^*_t \in \arg \min_{s \in S} \mathcal{L}^{(t)}(s)$$

For any measurable subset $A \subseteq S$, define the diameter $D(A) := \sup_{s,s' \in A} d(s,s')$ and volume $V_{\pi^{(0)}}(A) := \int_A \pi^{(0)}(s) \nu(ds) = \Pi^{(0)}(A)$.

Then we have the following bound on the regret for bounded losses:

**Lemma 1** (Hedge regret for bounded losses). The Hedge algorithm with non-increasing learning rates $(\eta_t)_t$ guarantees the following bound on the regret:

$$R^{(t)} \leq \frac{M^2}{8} \sum_{\tau=1}^{t} \eta_{\tau} + \left( \xi(\eta_t, L^{(t)}) - L^{(t)}(s^*_t) \right)$$

(2)
where
\[ \xi(\eta, f) = -\frac{1}{\eta} \log \int \pi^{(0)}(s) \exp(-\eta f(s)) \nu(ds) \]

Proof. We have
\[
\begin{align*}
\xi(\eta_{t+1}, L^{(t)}) - \xi(\eta_{t+1}, L^{(t+1)}) &= \frac{1}{\eta_{t+1}} \log \frac{\int \pi^{(0)}(s) \exp(-\eta_{t+1} \sum_{\tau=1}^{t+1} \ell(\tau)(s)) \nu(ds)}{\int \pi^{(0)}(u) \exp(-\eta_{t+1} \sum_{\tau=1}^{t} \ell(\tau)(u)) \nu(du)} \\
&= \frac{1}{\eta_{t+1}} \log \int \pi^{(t+1)}(s) \exp(-\eta_{t+1} \ell^{(t+1)}(s)) \nu(ds) \\
&= \frac{1}{\eta_{t+1}} \log E_{\pi^{(t+1)}}[\exp(-\eta_{t+1} \ell^{(t+1)})] \\
&\leq -E_{\pi^{(t+1)}}[\ell^{(t+1)}] + \frac{\eta_{t+1} M^2}{8}
\end{align*}
\]

where the last inequality follows from Hoeffding’s lemma. Summing the inequalities, we find that
\[
\sum_{\tau=1}^{t} \left( \xi(\eta_{\tau}, L^{(\tau-1)}) - \xi(\eta_{\tau}, L^{(\tau)}) \right) \leq -\sum_{\tau=1}^{t} E_{\pi^{(\tau)}}[\ell(\tau)] + \frac{M^2}{8} \sum_{\tau=1}^{t} \eta_{\tau}
\]

Rearranging, and observing that since \( L^{(0)} = 0, \xi(\eta, L^{(0)}) = \frac{1}{\eta} \log \int \pi^{(0)}(s) \nu(ds) = 0 \), we obtain
\[
\sum_{\tau=1}^{t} E_{\pi^{(\tau)}}[\ell(\tau)] \leq \frac{M^2}{8} \sum_{\tau=1}^{t} \eta_{\tau} + \sum_{\tau=1}^{t-1} \left( \xi(\eta_{\tau}, L^{(\tau)}) - \xi(\eta_{\tau+1}, L^{(\tau)}) \right) + \xi(\eta_{t}, L^{(t)})
\]

Next, we show that each term of the second sum in (3) is non-positive. Since \( \eta_{t+1} \leq \eta_{t} \) by assumption, it suffices to show that for any \( f \in \mathcal{B}, \eta \mapsto \xi(\eta, f) \) is increasing. Calculating the partial derivative w.r.t. \( \eta \) we have
\[
\frac{\partial \xi(\eta, f)}{\partial \eta} = \frac{1}{\eta^2} \log \int \pi^{(0)}(s) \exp(-\eta f(s)) \nu(ds) - \frac{1}{\eta} \int -f(s) \pi^{(0)}(s) \exp(-\eta f(s)) \nu(ds)
\]
\[
= -\frac{1}{\eta^2} \log \frac{1}{Z_f} - \frac{1}{\eta} \int -f(s) \pi_f(s) \nu(ds)
\]

where we use \( \pi_f(s) \) to denote the density function
\[
\pi_f(s) = \frac{\pi^{(0)}(s) \exp(-\eta f(s))}{Z_f}
\]

with \( Z_f \) the corresponding normalization constant. Then
\[
-\frac{\partial \xi(\eta, f)}{\partial \eta} = \frac{1}{\eta^2} \int \log \frac{1}{Z_f} \pi_f(s) \nu(ds) + \frac{1}{\eta^2} \int \log \exp(-\eta f(s)) \pi_f(s) \nu(ds)
\]
\[
= \frac{1}{\eta^2} \int \log \frac{\exp(-\eta f(s))}{Z_f} \pi_f(s) \nu(ds)
\]
\[
= \frac{1}{\eta^2} \int \log \frac{\pi_f(s)}{\pi^{(0)}(s)} \pi_f(s) \nu(ds)
\]
\[
= D_{KL}(\pi_f, \pi^{(0)})
\]
Therefore $-\frac{\partial \xi(\eta, f)}{\partial \eta}$ is proportional to the Kullback-Leibler divergence of $\pi_f$ with respect to $\pi^{(0)}$, and so $\eta \mapsto \xi(\eta, f)$ is non-increasing. The bound (3) then reduces to

$$
\sum_{\tau=1}^{t} \mathbb{E}_{\pi(\tau)}[\ell(\tau)] \leq \frac{M^2}{8} \sum_{\tau=1}^{t} \eta_{\tau} + \xi(\eta_{\tau}, L^{(t)})
$$

and we conclude by subtracting $L^{(t)}(s_0^*) = \min_{s \in S} \sum_{\tau=1}^{t} \ell(\tau)(s)$ from both sides.

We have two terms in the bound. The first term, $\frac{M^2}{8} \sum_{\tau=1}^{t} \eta_{\tau}$ is sublinear if the learning rates decay fast enough. In particular, if $\eta_{\tau} = \Theta(t^{-\alpha})$, $\alpha \in (0, 1)$, $\sum_{\tau=1}^{t} \eta_{\tau} = \mathcal{O} \left( \sum_{\tau=1}^{t} \int_{\tau}^{\tau+1} u^{-\alpha} du \right) = \mathcal{O}(t^{1-\alpha})$. To be able to guarantee a sublinear regret, we have to deal with the second term $\xi(\eta_{\tau}, L^{(t)} - L^{(t)}(s_0^*))$. We first specialize the bound to the case of Lipschitz losses.

Lemma 2 (Hedge regret for Lipschitz losses). Suppose that the loss functions $\ell(\tau)$ are L-Lipschitz uniformly in $t$. Consider a sequence of measurable subsets $\bar{S}_t \subset S$ with $s_0^* \in \bar{S}_t$ for all $t$. Then the Hedge algorithm with non-increasing learning rates $(\eta_{\tau})_t$ guarantees the following bound on the regret:

$$
R^{(t)} \leq \frac{M^2}{8} \sum_{\tau=1}^{t} \eta_{\tau} + tLD(\bar{S}_t) - \frac{\log V_{\pi^{(0)}}(\bar{S}_t)}{\eta_{\tau}}
$$

Proof. Since the loss functions $\ell(\tau)$ are uniformly L-Lipschitz, we have for all $s \in \bar{S}_t$ that $|\ell(\tau)(s) - \ell(\tau)(s_0^*)| \leq LD(s, s_0^*) \leq LD(\bar{S}_t)$. Hence $\ell(\tau)(s) \leq \ell(\tau)(s_0^*) + LD(\bar{S}_t)$, and $L^{(t)}(s) \leq L^{(t)}(s_0^*) + tLD(\bar{S}_t)$. Therefore

$$
\xi(\eta_{\tau}, L^{(t)}) = -\frac{1}{\eta_{\tau}} \log \int_{\bar{S}_t} \pi^{(0)}(s) \exp(-\eta_{\tau} L^{(t)}(s)) \nu(ds)
$$

$$
\leq -\frac{1}{\eta_{\tau}} \log \int_{\bar{S}_t} \pi^{(0)}(s) \exp(-\eta_{\tau} L^{(t)}(s_0^*)) \nu(ds)
$$

$$
\leq -\frac{1}{\eta_{\tau}} \log \int_{\bar{S}_t} \pi^{(0)}(s) \exp(-\eta_{\tau} (L^{(t)}(s_0^*) + tLD(\bar{S}_t))) \nu(ds)
$$

$$
= L^{(t)}(s_0^*) + tLD(\bar{S}_t) - \frac{1}{\eta_{\tau}} \log \int_{\bar{S}_t} \pi^{(0)}(s) \nu(ds)
$$

$$
= L^{(t)}(s_0^*) + tLD(\bar{S}_t) - \frac{1}{\eta_{\tau}} \log V_{\pi^{(0)}}(\bar{S}_t)
$$

Combining this inequality with Lemma 1 concludes the proof.

Lemma 2 provides regret bounds in terms of any sequence of subsets $(\bar{S}_t)_t$ that contain $s_0^*$, an optimal action in hindsight. However, this bound is only useful if one can construct such a sequence with appropriate relative decay rates of the diameters and volumes. Dividing the bound by $t$, we have

$$
\frac{R^{(t)}}{t} \leq \frac{M^2}{8} \frac{\sum_{\tau=1}^{t} \eta_{\tau}}{t} + LD(\bar{S}_t) - \frac{\log V_{\pi^{(0)}}(\bar{S}_t)}{t\eta_{\tau}}
$$

In particular, one needs the diameter of $\bar{S}_t$ to converge to zero, but the log volumes should converge to zero slower than $t\eta_{\tau}$. More precisely, we have the following corollaries.
Corollary 1. Consider the Hedge algorithm with learning rates \((\eta_t)\) such that \(\sum_{\tau=1}^{t} \eta_{\tau} = o(t)\). Suppose that there exists a sequence of subsets \((S_t)_{t} \) with \(s^*_t \in S_t \) for all \( t \), and such that \( D(S_t) = o(1) \) and \( \log V_{\pi(0)}(S_t) = o(\eta_t) \). Then the regret grows sublinearly, i.e.,

\[
\limsup_{t \to \infty} \frac{R_t}{t} \leq 0 \quad (7)
\]

Corollary 2. Consider the Hedge algorithm with constant learning rate \(\eta\). Suppose that for all \( t \), there exists a sequence of subsets \((\bar{S}_t)_{t} \) with \(s^*_t \in \bar{S}_t \), and such that \( D(\bar{S}_t) \to 0 \) and \( t^{-1} \log V_{\pi(0)}(\bar{S}_t) \to 0 \) as \( t \to \infty \). Then

\[
\limsup_{t \to \infty} \frac{R_t}{t} \leq \frac{M^2 \eta}{8} \quad (8)
\]

In the next Section, we give sufficient conditions on the action set \( S \) that guarantee the existence of a sequence \( \bar{S}_t \) with the desired properties.

3 Sublinear regret

We now restrict our attention to finite dimensional Euclidean spaces. We start with the simple case of convex \( S \), and construct a sequence \((\bar{S}_t) \) using a homothetic transformation centered at \( s^*_t \), similarly to Hazan et al. [2007].

3.1 On convex sets

Lemma 3. If \( S \subset \mathbb{R}^n \) is convex and \( \nu \) is the Lebesgue-uniform probability distribution over \( S \), then the sequence \( \bar{S}_t \) defined by the homothetic transformation

\[
\bar{S}_t = \{(1 - d_t)s^*_t + d_ts, s \in S\} = \{s^*_t + d_t(s - s^*_t), s \in S\} \quad (9)
\]

has diameters \( D(\bar{S}_t) = d_tD(S) \) and generalized volumes \( V_{\pi(0)}(\bar{S}_t) = \int_{\bar{S}_t} \pi(0)(s)ds = d^n_t \).

Proof. We first observe that

\[
D(\bar{S}_t) = \sup_{s,s' \in S_t} ||s - s'||
= \sup_{s,s' \in S} ||s^*_t + d_t(s - s^*_t) - s^*_t - d_t(s' - s^*)||
= d_tD(S)
\]

Moreover, if we denote by vol\((S)\) the Lebesgue-volume of \( S \), we have that

\[
V_{\pi(0)}(\bar{S}_t) = \int_{\bar{S}_t} \pi(0)(s)ds = \frac{1}{\text{vol}(S)} \int_{S_t} ds
= \frac{1}{\text{vol}(S)} \int_S |\det(d_tD_n)|ds' = d^n_t
\]

which proves the claim.
Corollary 3 (Hedge on convex compact subsets of $\mathbb{R}^n$). Let $S \subset \mathbb{R}^n$ be convex, and suppose that the $\ell^t$ are $L$-Lipschitz uniformly in time. If $\pi^{(0)}$ is the Lebesgue-uniform probability distribution over $S$, then for the Hedge algorithm with learning rates $\eta_t = \theta t^{-\alpha}$, $\alpha \in (0, 1)$, we have

$$\frac{R(t)}{t} \leq \frac{M^2 \theta}{8(1 - \alpha)} \frac{1}{t^{\alpha}} + \frac{L \text{D}(S)}{t} + \frac{n \log t}{\theta t^{1 - \alpha}}$$

(10)

In particular, the time-averaged regret is $\mathcal{O}\left(\frac{\log t}{\theta^\alpha}\right)$, where $\bar{\alpha} = \min(\alpha, 1 - \alpha)$.

Proof. Construct the sequence $\bar{S}_t$ as in Lemma 3, we have from the regret bound (6) that

$$\frac{R(t)}{t} \leq \frac{M^2 \theta}{8} \sum_{\tau=1}^{t} \tau^{-\alpha} + L \text{D}(S)d_t - \frac{n \log d_t}{\theta t^{1 - \alpha}}$$

Bounding $\sum_{\tau=1}^{t} \tau^{-\alpha} \leq \int_{0}^{t} \tau^{-\alpha} d\tau = \frac{1}{1 - \alpha} t^{1 - \alpha}$, we have

$$\frac{R(t)}{t} \leq \frac{M^2 \theta}{8(1 - \alpha)} \frac{1}{t^{\alpha}} + L \text{D}(S)d_t + \frac{n \log 1/d_t}{\theta t^{1 - \alpha}}$$

Taking $d_t = \frac{1}{t}$, we obtain (10).

□

Corollary 4 (Hedge on convex compact subsets of $\mathbb{R}^n$ with constant learning rate). Under the assumptions of Corollary 3, with a constant learning rate $\eta_t = \eta$, we have

$$\frac{R(t)}{t} \leq \frac{M^2 \eta}{8} \frac{1}{t^{\alpha}} + \frac{L \text{D}(S)}{t} + \frac{n \log t}{\eta t}$$

(11)

For a given horizon $T$, we can choose $\eta$ to minimize this bound. With $\eta = M^{-1} \sqrt{\frac{8n \log T}{T}}$, we have

$$\frac{R(T)}{T} \leq \frac{\text{D}(S)}{T} + M \sqrt{\frac{n \log T}{2T}}$$

(12)

3.2 On locally fat sets

The convexity assumption can, in fact, be relaxed to a local condition. Intuitively, to be able to use the sequence of sets $\bar{S}_t$ as constructed in Lemma 3, it suffices to find, for each $s^*_t$, a convex set $K_t$ containing $s^*_t$ such that its volume $V_\pi(s^*_t)$ is uniformly bounded below. This motivates the following definition:

Definition 1 (Local fatness). A compact set $S \subset \mathbb{R}^n$ is locally fat with respect to the measure $\pi$ if, for all $s \in S$, there exists a convex compact set $K_s \subseteq S$ such that $s \in K_s$ and $V_\pi(K_s) > 0$.

The maximal volume of such a set will be denoted

$$v_\pi(s) = \sup\{\text{vol}_\pi(K) : K \in K(S) \text{ and } s \in K\}$$

(13)

where $K(S)$ is the set of convex measurable subsets of $S$. Clearly, this implies that $v \leq 1$. See Fig. 1 (left) for an illustration of the local fatness condition.

Definition 2 (Uniform fatness). A compact set $S \subset \mathbb{R}^n$ is $v$-uniformly fat with respect to the distribution $\pi \in \Delta(S)$ if $\inf_{s \in S} v_\pi(s) \geq v$. 
3.2 On locally fat sets

Figure 1: Illustration of the local fatness condition for uniform \( \pi \). Left: \( K^*(s) \) is the maximizer of \( v_{\pi}(s) \) at \( s \), see (13).

**Lemma 4.** Suppose \( S \) is \( v \)-uniformly fat w.r.t. \( \pi^{(0)} \). Suppose that the loss functions are \( L \)-Lipschitz uniformly in time. Then the regret of the Hedge algorithm with uniform initial distribution and learning rates \( \eta_t = \theta t^{-\alpha} \), \( \alpha \in (0,1) \) satisfies the following bound

\[
\frac{R(t)}{t} \leq \frac{M^2 \theta}{8(1-\alpha)} \frac{1}{t^{\alpha}} + \frac{L D(S)}{t} + \frac{n \log t - \log v}{\theta t^{1-\alpha}}
\]

Proof. Since \( S \) is \( v \)-uniformly fat, for all \( t \), there exists a convex measurable subset \( K_{s_t^*} \subseteq S \) with \( s_t^* \in K_{s_t^*} \) and \( \text{vol}_{\pi^{(0)}}(K_{s_t^*}) \geq v \). Similarly to (9), define \( \tilde{S}_t \) as the homothetic transformation of \( K_{s_t^*} \) with center \( s_t^* \) and ratio \( d_t \). By Lemma 3, we have \( D(\tilde{S}_t) = d_t D(K_t^*) \leq d_t D(S) \) and

\[
V_{\pi^{(0)}}(\tilde{S}_t) = d_t^n V_{\pi^{(0)}}(K_t^*) \geq d_t^n v
\]

Applying the regret bound (6) with \( \eta_t = \theta t^{-\alpha} \), we have

\[
\frac{R(t)}{t} \leq \frac{M^2 \theta}{8(1-\alpha)} \frac{1}{t^{\alpha}} + \frac{L D(S)}{t} - \frac{\log V_{\pi^{(0)}}(\tilde{S}_t)}{\theta t^{1-\alpha}}
\]

\[
\leq \frac{M^2 \theta}{8(1-\alpha)} \frac{1}{t^{\alpha}} + d_t L D(S) - \frac{\log(d_t^n v)}{\theta t^{1-\alpha}}
\]

and we conclude by taking \( d_t = \frac{1}{t} \).

**Corollary 5.** Suppose \( S \) is \( v \)-uniformly fat, and that the loss functions are \( L \) Lipschitz uniformly in time. Then the Hedge algorithm with constant learning rate \( \eta \) satisfies

\[
\frac{R(t)}{t} \leq \frac{M^2 \eta}{8} + \frac{L D(S)}{t} + \frac{n \log t - \log v}{\eta t} \tag{14}
\]

In particular, for a known finite horizon \( T \), the constant learning rate which minimizes this bound is \( \eta_T = M^{-1} \sqrt{\frac{8(n \log T - \log v)}{T}} \), for which the regret is

\[
\frac{R^{(T)}}{T} \leq \frac{L D(S)}{T} + M \sqrt{\frac{n \log T - \log v}{2T}} 
\]

(15)

In particular, if \( S \) is convex, then \( \inf_{s \in S} v_{\pi}(S) = 1 \), and Corollary 4 becomes a special case of Corollary 5.
3.3 Doubling trick

Starting from a regret bound with constant learning rate, such as the bound of Corollary 5, one can design an algorithm with time-varying learning rates and sublinear regret, using what is sometimes called the doubling trick, see for example Cesa-Bianchi and Lugosi [2006]. We give the corresponding analysis in our setting.

Suppose the regret with constant learning rate $\eta$ satisfies the bound

$$R^{(t)} \leq \frac{M^2 t \eta}{8} + L D(S) + \frac{n \log t - \log v}{\eta}$$

Consider the Hedge algorithm with learning rates $\eta_t$ defined as follows: for all $t \in I_k = \{2^k, \ldots, 2^{k+1}-1\}$, $\eta_k$ is the learning rate which minimizes the regret with a horizon $T_k = 2^k$, the length of interval $I_k$. By Corollary 5, the regret on interval $I_k$, denoted $R^{(2^k)}$, is bounded as follows:

$$R^{(2^k)} \leq LD(S) + M \sqrt{\frac{2^{k+1} n k \log 2 - \log v}{2}} \leq LD(S) + M \sqrt{\frac{\max(n \log 2, - \log v)}{2}} \sqrt{(k+1)2^k}$$

Letting $C = M \sqrt{\frac{\max(n \log 2, - \log v)}{2}}$ and summing over intervals, we have for $m \geq 1$

$$\frac{1}{2^m} \sum_{k=0}^{m-1} R^{(2^k)} \leq \frac{mLD(S)}{2^m} + \frac{C}{2^m} \sum_{k=0}^{m-1} \sqrt{(k+1)2^k}$$

The first term is $O(m/2^m)$. To bound the second term, we can write

$$\sum_{k=0}^{m-1} \sqrt{(k+1)2^k} = \sum_{k=1}^{m} \sqrt{k2^{k-1}} \leq \sum_{k=1}^{m} k\sqrt{2^{(k-1)}}$$

where we bounded $k$ by $k$. Letting $S(x) = \sum_{k=1}^{\infty} x^k$, we have $\sum_{k=1}^{m} k\sqrt{2^{(k-1)}} = S'(\sqrt{2})$. We have

$$S(x) = \sum_{k=1}^{m} x^k = \frac{x^m - 1}{x - 1}$$

thus

$$S'(x) = \frac{(x-1)m x^{m-1} - (x^m - 1)}{(x-1)^2} = \frac{(m-1)x^m - mx^{m-1} + 1}{(x-1)^2} \leq \frac{mx^m + 1}{(x-1)^2}$$

which gives us the following bound:

$$\sum_{k=0}^{m-1} \sqrt{(k+1)2^k} \leq S'(\sqrt{2}) \leq \frac{m\sqrt{2}^m + 1}{(\sqrt{2} - 1)^2} \leq \frac{m\sqrt{2}^m + 1}{(\sqrt{2} - 1)^2}$$

Combining this with the previous inequality, we find that

$$\frac{1}{2^m} \sum_{k=0}^{m-1} R^{(2^k)} \leq \frac{mLD(S)}{2^m} + \frac{\sqrt{2}C}{(\sqrt{2} - 1)^2} \frac{m}{\sqrt{2}^m} \leq \frac{mLD(S)}{2^m} + M \frac{\sqrt{\max(n \log 2, - \log v)}}{(\sqrt{2} - 1)^2} \frac{m}{\sqrt{2}^m}$$
Thus,
\[ \frac{1}{2m} \sum_{k=0}^{m-1} R(2^k) = O \left( \frac{m}{\sqrt{2m}} \right) \]
In other words, \( \frac{R(t)}{t} = O \left( \frac{\log t}{\sqrt{t}} \right) \).

### 3.4 Logarithmic regret for exp concave functions

We now review the case when the loss functions are exp-concave, uniformly in time. That is, there exists \( \alpha > 0 \) such that for all \( t \), \( \exp(-\alpha \ell(t)) \) is concave.

**Lemma 5** (Hedge regret for exp-concave losses). Suppose that \( S \) is convex, and that the loss functions \( \ell(t) \) are \( \alpha \)-exp concave for some \( \alpha > 0 \). Then playing \( s(t) = E_{s \sim \pi(t)} s \) with \( \pi(t) \) given by the generalized Hedge algorithm with learning rate \( \eta = \alpha \) guarantees
\[ R(t) \leq n \log t + 1 \]

**Proof.** By exp-concavity, we first have a bound similar to Lemma 1, in terms of the potential \( \xi \):
\[ R(t) \leq \xi(\alpha, L(t)) - L(t)(s^*_t) \]
Indeed, we have
\[ \ell(t)(s_t) = -\frac{1}{\alpha} \log \left( \exp \left( -\alpha \ell(t) \left( E_{s \sim \pi(t)} s \right) \right) \right) \]
\[ \leq -\frac{1}{\alpha} \log \left( E_{s \sim \pi(t)} \exp \left( -\alpha \ell(t)(s) \right) \right) \]
\[ = \xi(\alpha, L(t)) - \xi(\alpha, L(t-1)) \]
where \( \xi(\alpha, f) = -\frac{1}{\alpha} \log \int \pi(0)(s) \exp(-\eta f(s)) \nu(ds) \), as defined in Lemma 1. Thus, summing over \( \tau \), we have
\[ \sum_{\tau=1}^{t} \ell(\tau)(s_{\tau}) \leq \xi(\alpha, L(t)) - \xi(\alpha, L(0)) = \xi(\alpha, L(t)) \]
and we conclude by subtracting \( L(t)(s^*_t) \) from both sides.

Therefore, to conclude, it suffices to show that \( \xi(\alpha, L(t)) - L(t)(s^*_t) \leq \frac{n(1+\log t)}{\alpha} \).

Now we construct the set \( \bar{S}_t \) using a homothetic transformation of \( S \) around \( s_t \) with radius \( d_t = \frac{1}{t} \), i.e.
\[ \bar{S}_t = \left\{ \left( 1 - \frac{1}{t} \right) s^*_t + \frac{1}{t} s, s \in S \right\} \]
Then by concavity of \( \exp -\alpha \ell(\tau) \), we have for all \( s \in \bar{S}_\tau \), we can write \( s = (1 - \frac{1}{t}) s^*_t + \frac{1}{t} s' \) for some \( s' \in S \), and
\[ \exp -\alpha \ell(\tau)(s) = \exp -\alpha \ell(\tau) \left( \left( 1 - \frac{1}{t} \right) s^*_t + \frac{1}{t} s' \right) \]
\[ \geq \left( 1 - \frac{1}{t} \right) \exp -\alpha \ell(\tau)(s^*_t) + \frac{1}{t} \exp -\alpha \ell(\tau)(s') \]
\[ \geq \left( 1 - \frac{1}{t} \right) \exp -\alpha \ell(\tau)(s^*_t) \]
Thus
\[
\xi(\alpha, L^{(t)}) = -\frac{1}{\eta} \log \int_S \pi^{(0)}(s) \exp(-\eta L^{(t)}(s)) \nu(ds)
\leq -\frac{1}{\eta} \log \int_{\tilde{S}_t} \pi^{(0)}(s) \prod_{t=1}^T \exp(-\eta \ell^{(t)}(s)) \nu(ds)
\leq -\frac{1}{\eta} \log \int_{\tilde{S}_t} \pi^{(0)}(s) \left(1 - \frac{1}{t}\right)^t \exp(-\eta L^{(t)}(s^*_t)) \nu(ds)
= L^{(t)}(s^*_t) - \frac{1}{\eta} \log V^{(0)}(\tilde{S}_t) - \frac{1}{\eta} t \log \left(1 - \frac{1}{t}\right)
\leq L^{(t)}(s^*_t) - \frac{1}{\eta} \log d^n + \frac{1}{\eta}
= L^{(t)}(s^*_t) + \frac{1 + n \log t}{\eta}
\]
which concludes the proof. \qed

4 Hedge on a finite cover

In this section, we compare our method to a related idea: for a given horizon, compute a finite cover of the set, such that the maximum difference of losses on each element of the cover is small enough, then perform a discrete learning algorithm on the finite cover.

More precisely, for a given accuracy \(\epsilon_T > 0\), assume that we can compute a finite cover \(A_T\) of \(S\), such that
\[
S \subseteq \bigcup_{\tilde{S}_T \in A_T} \tilde{S}_T
\]
and for all \(\tilde{S}_T \in A_T\),
\[
\sup_{s, s' \in \tilde{S}_T} |\ell^{(t)}(s) - \ell^{(t)}(s')| \leq \epsilon_T
\]
Since the loss functions are \(L\) Lipschitz, a sufficient condition is to have \(D(\tilde{S}_T) \leq \frac{\epsilon_T}{T}\). In \(\mathbb{R}^n\), the size of the cover is \(O\left(\frac{1}{\epsilon_T^n}\right)\).

If we call \(\tilde{R}^{(t)}\) the regret with respect to the discrete set, then running the discrete Hedge algorithm on the finite cover with learning rate \(\eta\) guarantees
\[
\frac{\tilde{R}^{(T)}}{T} \leq \frac{M^2}{8} \eta + \frac{\log |A|}{T \eta}
\]
and the optimal \(\eta\) given the horizon \(T\) is \(\eta_T = \frac{1}{2M} \sqrt{\frac{\log |A|}{2T}}\), for which the regret bound is
\[
\frac{\tilde{R}^{(T)}}{T} = O\left(\sqrt{\frac{\log |A|}{T}}\right)
\]
Since we incur at most \(\epsilon_T\) additional regret due to the variation of losses within each element of the cover, and since \(|A| = O\left(\frac{1}{\epsilon_T^n}\right)\) we have
\[
\frac{R^{(T)}}{T} = O\left(\sqrt{\frac{-n \log \epsilon_T}{T}} + \epsilon_T\right)
\]
and choosing $\epsilon_T = \frac{1}{\sqrt{T}}$, we obtain the following fixed horizon bound:

$$\frac{R^{(T)}}{T} = \mathcal{O}\left(\sqrt{\frac{n \log T}{T}}\right)$$

which matches the bound of Corollary 5. To generalize the method to time-varying learning rates, one could define a hierarchical sequence of covers $A_t$, and run the Hedge algorithm at different levels of the hierarchy on epochs of doubling length.

The Hedge algorithm on uniformly sets is conceptually similar to the idea of working with a finite cover. This is most visible in the proof of Lemma 4, where we rely on the construction of a set around $s_t^*$ with small enough diameter (and large enough volume). However, to apply this algorithm, one does not need to explicitly construct a cover, which could, in itself, be hard.

5 Simulations

We illustrate these results with some numerical simulations in $\mathbb{R}^2$.

We consider a first example where the feasible set is the union of rectangles and the loss functions are convex quadratic, of the form $\ell(t)(s) = \langle s, Q_t s \rangle + \langle a_t, s \rangle + c_t$ with $Q_t$ positive semi-definite, $a_t \in \mathbb{R}^2$ and $c_t \in \mathbb{R}$. In this case, starting from the uniform distribution, the probability densities generated by the Hedge algorithm are Gaussian, since

$$\pi^{(t+1)}(s) \propto \exp\left(-\eta_{t+1} \sum_{\tau=1}^t \ell^{(\tau)}(s)\right)$$

$$\propto \exp\left(-\eta_{t+1} \left( s_t \sum_{\tau=1}^t Q_\tau s + \sum_{\tau=1}^t a_\tau, s \right)\right)$$

$$\propto \exp\left(-\eta_{t+1} \left( s - s_t, \sum_{\tau=1}^t Q_\tau (s - s_t) \right)\right)$$

where $s_t \sum_{\tau=1}^t Q_\tau = \frac{1}{2} \sum_{\tau=1}^t a_\tau$.

Therefore if we have the loss function at each iteration, we can explicitly maintain the Gaussian distribution generated by the Hedge algorithm, and efficiently sample from it. Running the simulation for different choices of learning rates yields the results in Figure 2, where the average cumulative regret $R^{(t)}_t$ is given in logarithmic scale.

We also run simulations with general loss functions (in our particular implementation, we chose to encode the loss as the negative exponential of a quadratic). In this case, the Hedge distributions do not have any remarkable form, and sampling from these distributions can be done using general sampling methods, such as importance sampling or Metropolis Hastings. The results are reported in Figure 3, where the average cumulative regret $R^{(t)}_t$ is given in logarithmic scale, together with the corresponding upper bound predicted by our results. In these simulations (quadratic and exponential losses), the average regret converges to zero, and the decay rate is consistent with our results. In particular, the logarithmic plot ($\log R^{(t)}_t, \log t$) has a linear asymptote, with a slope which is consistent with the convergence rates of Lemma 4. In this particular instance, the
actual convergence rate seems to be faster than the bounds of the Lemma. The bound predicts a convergence rate of $O(t^{-\min(\alpha,1-\alpha)})$ with a learning rate sequence $t^{-\alpha}$, which is optimal for $\alpha = \frac{1}{2}$, while the plots show faster convergence for learning rates that have a slower decay.

Figure 3: Cumulative regret for quadratic losses, averaged across multiple simulations.
6  Dual averaging

In this last section, we explore possible generalization of the Hedge algorithm. The sequence of density functions generated by the Hedge algorithm, given by equation (1), can be viewed as a particular instance of the dual averaging method.

The dual averaging method Nesterov [2009] is a general method for solving constrained optimization problems. Given a sequence of vectors \( g(t) \) (in the context of convex optimization these are usually taken to be sub gradient vectors of the convex function to minimize), the method projects, at each step, the cumulative vector \( z(t) = \sum_{\tau=1}^{t} g(\tau) \) on the feasible set, using a Bregman projection. The method is summarized in Algorithm 2.

### Algorithm 2

Dual averaging method with input vectors \( (g(t)) \) and learning rates \( (\eta_t) \)

1: for \( t \in \mathbb{N} \) do
2: \( z(t) = \sum_{\tau=1}^{t} g(\tau) \)
3: \( x(t+1) = \arg \min_{x \in X} \langle x, z(t) \rangle + \frac{1}{\eta_{t+1}} \psi(x) \)
4: end for

Here \( \psi \) is assumed to be \( \ell \)-strictly convex w.r.t. some norm \( \| \cdot \| \). To give a connection with mirror descent, if \( \eta_{t+1} z(t) = -\nabla \psi(w(t)) \), then the minimization step becomes

\[
x(t+1) = \arg \min_{x \in X} \left( x, z(t) \right) + \frac{1}{\eta_{t+1}} \psi(x)
= \arg \min_{x \in X} \left( x - w(t), \nabla \psi(w(t)) \right) + \psi(x) - \psi(w(t))
= \arg \min_{x \in X} D_\psi(x, w(t))
\]

so the dual averaging method projects the vector \( w(t) = (\nabla \psi)^{-1} \left( -\eta_{t+1} \sum_{\tau=1}^{t} g(\tau) \right) \) on the feasible set, using the Bregman divergence \( D_\psi \).

In particular, consider the case where \( g(t) \) are bounded functions on \( S \), and \( X \) is the set of distributions over the set \( S \), that are absolutely continuous with respect to a reference distribution \( \nu \). The set \( X \) is convex, and we can define the inner product as follows

\[
\langle x(t), g(t) \rangle = \int_{s \sim x(t)} g(t)(s) \nu(ds)
\]

Then the cumulative expected regret is

\[
R(t) = \sum_{\tau=1}^{t} \langle x(\tau), g(\tau) \rangle - \inf_{s \in S} \sum_{\tau=1}^{t} g(\tau)(s)
= \sum_{\tau=1}^{t} \langle x(\tau), g(\tau) \rangle - \inf_{x \in X} \left( x, \sum_{\tau=1}^{t} g(\tau)(s) \right)
\]

We next show that the dual averaging method achieves a sublinear regret under appropriate assumptions on the distance generating function \( \psi \).

The following Lemma will be useful in proving the main result.
Lemma 6. Suppose $\psi$ is $\ell$-strongly convex w.r.t. $\|\cdot\|$. Let $z_1, z_2 \in \mathbb{R}^d$, and $x_i = \arg \min_{x \in X} \langle z_i, x \rangle + \frac{1}{\eta} \psi(x)$. Then
\[ \|x_1 - x_2\| \leq \eta \frac{\ell}{\eta} \|z_1 - z_2\|_*. \]

Proof. By optimality of $x_i$, we must have
\[ \left\langle y - x_i, z_i + \frac{1}{\eta} \nabla \psi(x_i) \right\rangle \geq 0 \quad \forall y \in \mathcal{X}. \]

Thus
\[ \left\langle x_2 - x_1, z_1 + \frac{1}{\eta} \nabla \psi(x_1) \right\rangle \geq 0 \]
\[ \left\langle x_1 - x_2, z_2 + \frac{1}{\eta} \nabla \psi(x_2) \right\rangle \geq 0 \]

Thus
\[ \langle x_2 - x_1, z_1 - z_2 \rangle \geq \frac{1}{\eta} \langle x_1 - x_2, \nabla \psi(x_1) - \nabla \psi(x_2) \rangle \]

Therefore
\[ \|z_1 - z_2\|_* \|x_1 - x_2\| \geq \langle x_2 - x_1, z_1 - z_2 \rangle \quad \text{by Hölder's inequality} \]
\[ \geq \frac{1}{\eta} \langle x_1 - x_2, \nabla \psi(x_1) - \nabla \psi(x_2) \rangle \quad \text{by optimality} \]
\[ \geq \frac{\ell}{\eta} \|x_1 - x_2\|^2 \quad \text{by strong convexity of } \psi \]

We now show a general bound for any sequence of vectors $g^{(t)}$

Lemma 7 (Dual averaging bound). Suppose that $\psi$ is $\ell_\psi$-strongly convex w.r.t. a norm $\|\cdot\|$, and that $(\eta_t)$ is decreasing. Then for any sequence $(g^{(t)})_t \in \mathbb{R}^d$, and any $x \in \mathcal{X}$, the dual averaging method guarantees
\[ \sum_{\tau \leq t} \left\langle x^{(\tau)} - x, g^{(\tau)} \right\rangle \leq \frac{1}{\eta_{t+1}} \psi(x) - \frac{1}{\eta_1} \psi(x^{(1)}) + \frac{1}{\ell} \sum_{\tau=1}^{t} \eta_{\tau+1} \|g^{(\tau)}\|_*^2 + \sum_{\tau=1}^{t-1} \psi(x^{(\tau+1)}) \left( \frac{1}{\eta_{\tau}} - \frac{1}{\eta_{\tau+1}} \right) \]

In particular, with a constant learning rate $\eta$, the sum is zero and the bound reduces to
\[ \sum_{\tau \leq t} \left\langle x^{(\tau)} - x, g^{(\tau)} \right\rangle \leq \frac{1}{\eta} \left( \psi(x) - \psi(x^{(1)}) \right) + \frac{\eta}{\ell} \sum_{\tau \leq t} \|g^{(\tau)}\|_*^2 \]

and if $\psi$ is non-negative, with any decreasing sequence of learning rates, the sum is non-positive, and the bound reduces to
\[ \sum_{\tau \leq t} \left\langle x^{(\tau)} - x, g^{(\tau)} \right\rangle \leq \frac{1}{\eta_{t+1}} \psi(x) + \frac{1}{\ell} \sum_{\tau=1}^{t} \eta_{\tau+1} \|g^{(\tau)}\|_*^2 \]

Proof. We have
\[ \sum_{\tau=1}^{t} \left\langle x^{(\tau)} - x, g^{(\tau)} \right\rangle = \sum_{\tau=1}^{t} \left( \left\langle x^{(\tau)} - x^{(\tau+1)}, g^{(\tau)} \right\rangle + \left\langle x^{(\tau+1)} - x, g^{(\tau)} \right\rangle \right) \]

and we bound each term.
First term  We first rewrite the sum in terms of $z(t)$
\[
\sum_{\tau=1}^{t} \langle x^{(\tau+1)} - x, g^{(\tau)} \rangle = \sum_{\tau=1}^{t} \langle x^{(\tau+1)}, g^{(\tau)} \rangle - \langle x, z(t) \rangle \\
= \sum_{\tau=1}^{t} \langle x^{(\tau+1)}, z^{(\tau)} - z^{(\tau-1)} \rangle - \langle x, z(t) \rangle \\
= \sum_{\tau=1}^{t} \langle x^{(\tau+1)}, z^{(\tau)} \rangle - \sum_{\tau=0}^{t-1} \langle x^{(\tau+2)}, z^{(\tau)} \rangle - \langle x, z(t) \rangle \\
= \sum_{\tau=1}^{t-1} \langle x^{(\tau+1)} - x^{(\tau+2)}, z^{(\tau)} \rangle + \langle x^{(t+1)}, z(t) \rangle - \langle x, z(t) \rangle \\
= \sum_{\tau=1}^{t-1} \langle x^{(\tau+1)} - x^{(\tau+2)}, z^{(\tau)} \rangle + \langle x^{(t+1)} - x, z(t) \rangle
\]

Then since $x^{(\tau+1)}$ is the minimizer of $\langle x, z^{(\tau)} \rangle + \frac{1}{\eta_{t+1}} \psi(x)$, we have
\[
\langle x^{(\tau+1)} - x^{(\tau+2)}, z^{(\tau)} \rangle \leq \frac{1}{\eta_{t+1}} \left( \psi(x^{(\tau+2)}) - \psi(x^{(\tau+1)}) \right)
\]

Similarly,
\[
\langle x^{(t+1)} - x, z(t) \rangle \leq \frac{1}{\eta_{t+1}} \left( \psi(x) - \psi(x^{(t+1)}) \right)
\]

Summing the inequalities, we have
\[
\sum_{\tau=1}^{t} \langle x^{(\tau+1)} - x, g^{(\tau)} \rangle \leq \sum_{\tau=1}^{t-1} \langle x^{(\tau+1)} - x^{(\tau+2)}, z^{(\tau)} \rangle + \langle x^{(t+1)} - x, z(t) \rangle \\
\leq \sum_{\tau=1}^{t-1} \frac{1}{\eta_{t+1}} \left( \psi(x^{(\tau+2)}) - \psi(x^{(\tau+1)}) \right) + \frac{1}{\eta_{t+1}} \left( \psi(x) - \psi(x^{(t+1)}) \right) \\
= \frac{1}{\eta_{t+1}} \psi(x) + \sum_{\tau=2}^{t-1} \psi(x^{(\tau+1)}) \left( \frac{1}{\eta_{\tau}} - \frac{1}{\eta_{t+1}} \right) - \frac{1}{\eta_{2}} \psi(x^{(2)})
\]

Second term  We can bound the second term as follows
\[
\langle x^{(\tau)} - x^{(\tau+1)}, g^{(\tau)} \rangle \leq \|g^{(\tau)}\|_* \|x^{(\tau)} - x^{(\tau+1)}\| \\
\leq \frac{\eta_{t+1}}{L} \|g^{(\tau)}\|_* \|z^{(\tau-1)} - z^{(\tau)}\|_* \\
= \frac{\eta_{t+1}}{L} \|g^{(\tau)}\|_*^2
\]

by Hölder

by Lemma 6

We conclude by summing the two inequalities.

Corollary 6. Let $\psi$ be a non-negative, strongly convex functions, $(\eta_t)$ a decreasing sequence of learning rates, and suppose that the sequence $g^{(\tau)}$ is uniformly bounded in the dual norm, i.e. there
exists $G > 0$ such that for all $t$, $\|g(t)\|_* \leq G$. Then

$$\sum_{\tau \leq t} \left\langle x(\tau) - x, g(\tau) \right\rangle \leq \frac{1}{\eta t + 1} \psi(x) + \frac{1}{\ell} \sum_{\tau = 1}^{t} \eta_{\tau + 1}$$

In particular, the regret is sublinear for $\eta_t = \Theta(t^{-\alpha})$, $\alpha \in (0, 1)$.

7 Conclusion

We studied an online optimization problem over a general subset $S$ of $\mathbb{R}^n$. Previous work shows that for the case in which $S$ is convex, one can achieve a $O(\sqrt{t})$ regret using a gradient descent method when the losses are convex, and $O(\log t)$ regret using a generalized Hedge algorithm when the losses are exp-concave. We consider Lipschitz losses, and relax the convexity assumption of $S$. In particular, we show that as long as the set is uniformly fat, i.e. one can always fit a convex set of minimal volume $v$ around any point of the set, then the generalized Hedge algorithm achieves a $O(\sqrt{t \log t})$ regret. The analysis is conceptually similar to the idea of working with a finite cover of the set, and achieves the same asymptotic bound. However, this method does not require explicitly computing a cover, and is potentially more efficient, depending on how easy it is to sample from the distribution generated by the Hedge algorithm. A question which remains open is whether the uniform fatness condition is necessary in some sense. If the set is not uniformly fat, then for any given volume $v$, there exists a point $s$ in $S$ such that no convex set of volume $v$ exists around $s$, and one could construct a particular sequence of loss functions such that the sequence of optima $s^*_t$ has vanishing $v(s^*_t)$, and use it to construct a lower bound on the regret for algorithms which, loosely speaking, ‘cannot shift probability mass too fast’, e.g. algorithms with an upper bound on the K-L divergence between $\pi(t)$ and $\pi(t+1)$ (this is true for the Hedge algorithm).

Finally, in our analysis, we did not attempt to optimize over the dependence on the dimension $n$ of the space. In particular, the bound of Corollary 4 has $\sqrt{n}$ term. A more careful analysis may reveal that one can reduce this dependence.
References


James Hannan. Approximation to Bayes risk in repeated plays. Contributions to the Theory of Games, 3:97–139, 1957. 2


