

Homework 1: Online Learning Algorithms and Games

Instructor: Mengxiao Zhang

Due: Oct. 18th

Total points: 50, plus 10 bonus points. Upload the PDF generated by LaTeX. Unless otherwise stated, all logarithms are natural logarithms and $N \geq 2$.

1. Doubling Trick

(6 pts)

We have seen that, for any fixed horizon T , Hedge with learning rate $\eta = \sqrt{(\ln N)/T}$ satisfies

$$\text{Reg}_T \leq 2\sqrt{T \ln N}.$$

This tuning is unavailable when the time horizon T is unknown. To handle this, consider the following anytime version of Hedge. The vector $\tilde{L}$ stores cumulative losses only within the current restart.

Algorithm 1: Hedge with the doubling trick

```

Initialize  $\tilde{L}_i = 0$  for every  $i \in [N]$ ,  $T_0 = 1$ , and  $\eta = \sqrt{(\ln N)/T_0}$ ;
for  $t = 1, 2, \dots$  do
    if  $t \geq 2T_0$  then
        set  $T_0 \leftarrow 2T_0$ ;
        reset  $\tilde{L}_i \leftarrow 0$  for every  $i \in [N]$  and set  $\eta \leftarrow \sqrt{(\ln N)/T_0}$ ;
    set  $p_{t,i} = \frac{\exp(-\eta \tilde{L}_i)}{\sum_{j=1}^N \exp(-\eta \tilde{L}_j)}$  for every  $i \in [N]$ ;
    play  $p_t$  and observe  $\ell_t \in [0, 1]^N$ ;
    update  $\tilde{L}_i \leftarrow \tilde{L}_i + \ell_{t,i}$  for every  $i \in [N]$ ;

```

Prove that Algorithm 1 guarantees $\text{Reg}_T \leq \mathcal{O}(\sqrt{T \ln N})$ simultaneously for every $T \geq 1$.

2. Online Mirror Descent

(16 pts)

In the lectures, we discussed FTRL. In this problem, you will analyze a very similar (and popular) online learning algorithm called Online Mirror Descent (OMD). Let $\Omega \subseteq \mathbb{R}^d$ be nonempty, closed, and convex. Let $\psi : \Omega \rightarrow \mathbb{R}$ be continuous and convex, and assume that ψ is differentiable at every iterate. For $u \in \Omega$ and $w \in \Omega$ at which ψ is differentiable, define its Bregman divergence by

$$D_\psi(u, w) \triangleq \psi(u) - \psi(w) - \langle \nabla \psi(w), u - w \rangle.$$

Initialize $w_1 \in \Omega$. For $t = 1, \dots, T$, OMD chooses w_{t+1} as follows:

$$w_{t+1} \in \operatorname{argmin}_{w \in \Omega} \left\{ \langle w, g_t \rangle + \frac{1}{\eta} D_\psi(w, w_t) \right\}, \quad (1)$$

where each $f_t : \Omega \rightarrow \mathbb{R}$ is convex, $g_t \in \partial f_t(w_t)$, and $\eta > 0$. Assume that each minimum is attained.

- (a) **(4 pts)** Use the first-order optimality condition for (1) to prove that, for every $u \in \Omega$,

$$\eta \langle w_{t+1} - u, g_t \rangle \leq D_\psi(u, w_t) - D_\psi(u, w_{t+1}) - D_\psi(w_{t+1}, w_t). \quad (2)$$

Use the above result to show that

$$\sum_{t=1}^T \langle w_t - u, g_t \rangle \leq \frac{D_\psi(u, w_1)}{\eta} + \sum_{t=1}^T \langle w_t - w_{t+1}, g_t \rangle - \frac{1}{\eta} \sum_{t=1}^T D_\psi(w_{t+1}, w_t). \quad (3)$$

- (b) **(4 pts)** Suppose that ψ is 1-strongly convex with respect to a norm $\|\cdot\|$. Prove the following “stability” bound for w_t :

$$\|w_t - w_{t+1}\| \leq \eta \|g_t\|_*.$$

Use this fact in (3) to conclude that

$$\sum_{t=1}^T \langle w_t - u, g_t \rangle \leq \frac{D_\psi(u, w_1)}{\eta} + \eta \sum_{t=1}^T \|g_t\|_*^2. \quad (4)$$

- (c) **(4 pts)** Take $\Omega = \Delta(N)$, initialize $w_1 = \frac{1}{N} \mathbf{1}$, and let $g_t = \ell_t \in [0, 1]^N$ and

$$\psi(p) = \sum_{i=1}^N p_i \ln p_i, \quad 0 \ln 0 \triangleq 0.$$

Here ψ is differentiable on the relative interior of $\Delta(N)$ and is extended continuously to its boundary. Show that the OMD iterates have positive coordinates and that the update is exactly Hedge. Then use (4) to recover an $\mathcal{O}(\sqrt{T \ln N})$ regret guarantee. (Hint: You may use Lagrange multipliers and the fact that negative entropy is 1-strongly convex with respect to $\|\cdot\|_1$.)

- (d) **(4 pts)** Show that OMD with $\psi(w) = \frac{1}{2} \|w\|_2^2$ becomes projected Online Gradient Descent introduced in class,

$$w_{t+1} = \Pi_\Omega(w_t - \eta g_t).$$

Suppose $\|g_t\|_2 \leq G$ for $G > 0$ and $\text{diam}(\Omega) \triangleq \max_{w, u \in \Omega} \|w - u\|_2 \leq D$ for $D > 0$. Show that with an appropriate choice of $\eta > 0$, OMD achieves regret at most $\mathcal{O}(DG\sqrt{T})$ against every $u \in \Omega$. (Hint: compute $D_\psi(x, y)$ with $\psi(x) = \frac{1}{2} \|x\|_2^2$.)

3. Data-Dependent Guarantees: Hedge versus Prod

(9 pts)

In the lecture, we mentioned that Prod’s comparator second-order guarantee can be sharper than Hedge’s first-order guarantee. You will compare these guarantees in this problem. Let $i^* \in \arg\min_{i \in [N]} L_{T,i}$, and define

$$L_T^* \triangleq L_{T,i^*} = \sum_{t=1}^T \ell_{t,i^*}, \quad V_T^* \triangleq \sum_{t=1}^T \ell_{t,i^*}^2,$$

where $\ell_t \in [0, 1]^N$. You may use the following fixed-learning-rate guarantees from lecture: for Hedge, we have

$$\text{Reg}_T \leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^N p_{t,i} \ell_{t,i}^2, \quad (5)$$

and for Prod, we have for all $i \in [N]$

$$\sum_{t=1}^T \langle p_t, \ell_t \rangle - L_{T,i} \leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \ell_{t,i}^2. \quad (6)$$

where Prod requires $0 < \eta \leq 1/2$.

- (a) **(4 pts)** Prove that, for every $0 < \eta \leq 1/2$, $\text{Reg}_T^{\text{Hedge}} \leq \frac{2 \ln N}{\eta} + 2\eta L_T^*$. Then, show that when $\eta_H = \min \left\{ \frac{1}{2}, \sqrt{\frac{\ln N}{L_T^* + 1}} \right\}$, Hedge has the following guarantee

$$\text{Reg}_T^{\text{Hedge}} \leq \mathcal{O} \left(\sqrt{(L_T^* + 1) \ln N} + \ln N \right).$$

When the loss of the best action is small, this guarantee can be sharper than the classic $\mathcal{O}(\sqrt{T \ln N})$ regret bound.

- (b) **(3 pts)** Apply (6) to i^* and set $\eta_P = \min \left\{ \frac{1}{2}, \sqrt{\frac{\ln N}{V_T^* + 1}} \right\}$. Prove the comparator second-order guarantee

$$\text{Reg}_T^{\text{Prod}} \leq \mathcal{O} \left(\sqrt{(V_T^* + 1) \ln N} + \ln N \right).$$

Explain why this always implies the bound of the same order as Hedge's as shown in part (a). In addition, construct a specific environment $\{\ell_t\}_{t \in [T]}$ for which the Prod bound in this part is $\mathcal{O}(\ln N)$, whereas the expression in the bound in part (a) grows on the order of $T^{1/4} \sqrt{\ln N} + \ln N$, giving a worse dependence on T for fixed N .

- (c) **(2 pts)** Neither L_T^* nor V_T^* is known before the algorithm starts. Describe a doubling scheme similar to Question 1 that recovers the Hedge bound in part (a) without prior knowledge of L_T^* , and justify its guarantee. Explain why the same argument does not immediately recover the Prod bound in part (b). (Hint: the cumulative squared loss of the best expert in hindsight through round t need not be nondecreasing in t .)

4. From Constant Total Regret to Constant Individual Regret (6 pts)

Consider repeated play of a two-player zero-sum game with payoff matrix $A \in [-1, 1]^{M \times N}$, where $M, N \geq 2$. At round t , the row player chooses $x_t \in \Delta(M)$ and minimizes $x_t^\top A y_t$, while the column player chooses $y_t \in \Delta(N)$ and maximizes the same quantity. Define their regrets as follows:

$$\begin{aligned} \text{Reg}_T^X &\triangleq \sum_{t=1}^T x_t^\top A y_t - \min_{x \in \Delta(M)} \sum_{t=1}^T x^\top A y_t, \\ \text{Reg}_T^Y &\triangleq \max_{y \in \Delta(N)} \sum_{t=1}^T x_t^\top A y - \sum_{t=1}^T x_t^\top A y_t. \end{aligned}$$

Suppose both players run Optimistic Hedge from the uniform distribution with the same fixed learning rate $\eta > 0$, using their previous-round loss vectors as predictions, with $\ell_t^X = A y_t$, $\ell_t^Y = -A^\top x_t$, and $\ell_0^X = \ell_0^Y = 0$. In the lecture, we showed that for $\eta = \frac{1}{4}$, $\text{Reg}_T^X + \text{Reg}_T^Y \leq \mathcal{O}(\ln(MN))$. However, this does not immediately imply the individual bounds $\text{Reg}_T^X \leq \mathcal{O}(\ln(MN))$ and $\text{Reg}_T^Y \leq \mathcal{O}(\ln(MN))$.

- (a) **(2 pts)** Show a concrete example of $\{(x_t, y_t)\}_{t \in [T]}$ with $T = 2$ and a 2×2 matrix A such that $\text{Reg}_T^X < 0$. You do not need to provide $\{(x_t, y_t)\}_{t \in [T]}$ that are generated by running Optimistic Hedge.
- (b) **(2 pts)** Define the second-order path lengths

$$S_x \triangleq \sum_{t=2}^T \|x_t - x_{t-1}\|_1^2, \quad S_y \triangleq \sum_{t=2}^T \|y_t - y_{t-1}\|_1^2.$$

Show that as long as $0 < \eta \leq \frac{1}{4}$, we have $S_x + S_y \leq \frac{2 \ln(MN)}{\eta} + 4\eta$. (Hint: use the regret inequality from lecture, and recall the definition of saddle-point gap.)

- (c) **(2 pts)** Use the above results to prove that $\text{Reg}_T^X \leq \mathcal{O}(\ln(MN))$ and $\text{Reg}_T^Y \leq \mathcal{O}(\ln(MN))$ for $\eta = \frac{1}{4}$. Thus each player has an individual regret upper bound independent of T , not merely a bound on total regret.

5. Average-iterate versus Last-iterate Convergence

(13 pts)

Return to a two-player zero-sum game in which the row player minimizes $x^\top A y$ and the column player maximizes it. Define the average iterates and the saddle-point gap by

$$\bar{x}_t \triangleq \frac{1}{t} \sum_{s=1}^t x_s, \quad \bar{y}_t \triangleq \frac{1}{t} \sum_{s=1}^t y_s,$$

$$\text{Gap}(x, y) \triangleq \max_{v \in \Delta(N)} x^\top A v - \min_{u \in \Delta(M)} u^\top A y.$$

In the lecture, we showed that when both players run Optimistic Hedge with $\eta = 1/4$, the average iterates satisfy $\text{Gap}(\bar{x}_T, \bar{y}_T) \leq \mathcal{O}(\ln(MN)/T)$. This does not by itself imply that $\text{Gap}(x_T, y_T) \rightarrow 0$, which is last-iterate convergence in saddle-point gap. In this problem, you will compare these two notions of convergence empirically for four online learning algorithms.

Consider the game $A = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}$. Its unique Nash equilibrium is $x^* = y^* = (1/3, 2/3)$. Use $T = 10^4$ and initialize $x_1 = y_1 = (1/2, 1/2)$ in every run.

- (a) **(5 pts)** Implement Hedge, Optimistic Hedge, Euclidean FTRL, and Optimistic Euclidean FTRL. In each run, both players use the same algorithm and the same fixed learning rate. Test each $\eta \in \{0.125, 0.25, \frac{1}{\sqrt{T}}\}$. At round t , compute the exact loss vectors $\ell_t^X = A y_t$ and $\ell_t^Y = -A^\top x_t$, then update both players simultaneously.

For a generic player with distribution $p_t \in \Delta(d)$ and loss vector ℓ_t , use the following FTRL update rules:

$$\begin{aligned}
\text{Hedge:} \quad p_{t+1,i} &\propto \exp \left(-\eta \sum_{s=1}^t \ell_{s,i} \right), \\
\text{Optimistic Hedge:} \quad p_{t+1,i} &\propto \exp \left(-\eta \left(\sum_{s=1}^t \ell_{s,i} + \ell_{t,i} \right) \right), \\
\text{Euclidean FTRL:} \quad p_{t+1} &= \Pi_{\Delta(d)} \left(p_1 - \eta \sum_{s=1}^t \ell_s \right), \\
\text{Optimistic Euclidean FTRL:} \quad p_{t+1} &= \Pi_{\Delta(d)} \left(p_1 - \eta \left(\sum_{s=1}^t \ell_s + \ell_t \right) \right).
\end{aligned}$$

Here $\Pi_{\Delta(d)}$ denotes Euclidean projection onto the simplex, and the proportional updates are normalized to sum to one. For the optimistic variants, the prediction for round $t+1$ is ℓ_t , with $\ell_0 = 0$.

- (b) **(3 pts)** For every algorithm and learning rate, plot $\text{Gap}(x_t, y_t)$ and $\text{Gap}(\bar{x}_t, \bar{y}_t)$ against t . Investigate empirically whether the algorithms converge to equilibrium and, if so, how fast. Choose appropriate axes to reveal the convergence behavior and support your conclusions. For plotting only, you can replace gaps below 10^{-12} by 10^{-12} and state this convention in the captions. For this game, the gap can be computed as $\text{Gap}(x, y) = \max_j (A^\top x)_j - \min_i (Ay)_i$. Also plot the strategy trajectories $(x_{t,1}, y_{t,1})$ for each choice of η , marking the equilibrium $(1/3, 1/3)$.
- (c) **(5 pts)** Discuss your observations. Which algorithms and learning rates appear to achieve last-iterate convergence? Do any runs exhibit decreasing average-iterate gaps while the current iterates continue to oscillate?

6. (10 bonus pts) Extension to General-Sum Games.

A two-player general-sum game is defined by two payoff matrices $U_1, U_2 \in [-1, 1]^{M \times N}$. At round t , the players choose $x_t \in \Delta(M)$ and $y_t \in \Delta(N)$ and receive payoffs $x_t^\top U_1 y_t$ and $x_t^\top U_2 y_t$, respectively. Each player maximizes her own payoff; we do not require $U_2 = -U_1$ (as in a two-player zero-sum game). Their external regrets are

$$\begin{aligned} \text{Reg}_T^X &\triangleq \max_{x \in \Delta(M)} \sum_{t=1}^T x^\top U_1 y_t - \sum_{t=1}^T x_t^\top U_1 y_t, \\ \text{Reg}_T^Y &\triangleq \max_{y \in \Delta(N)} \sum_{t=1}^T x_t^\top U_2 y - \sum_{t=1}^T x_t^\top U_2 y_t. \end{aligned}$$

Suppose both players run Optimistic Hedge from the uniform distribution with the same fixed learning rate $\eta > 0$. Their loss vectors are $\ell_t^X = -U_1 y_t$ and $\ell_t^Y = -U_2^\top x_t$. Each player uses her previous-round loss vector as the optimism, with $\ell_0^X = \ell_0^Y = 0$.

Try to extend the argument in Question 4 to this setting. Can that argument, with a suitable choice of η , possibly depending on M, N , and T , establish

$$\text{Reg}_T^X \leq \mathcal{O}(\ln(MN)), \quad \text{Reg}_T^Y \leq \mathcal{O}(\ln(MN))$$

for every pair of payoff matrices and every horizon T , with a universal constant in the bounds? If so, specify the learning rate and prove the bounds rigorously. If the argument does not go through even with two players, identify a concrete obstruction and explain what additional inequality or assumption would allow you to complete it. Full credit can be earned by either proving the stated bounds or identifying and justifying such an obstruction in the two-player setting.

Optional: If you are interested, consider whether the argument extends to games with more than two players.