Homework 2

Instructor: Haipeng Luo

1. (**No-regret v.s Best Response**) In Lecture 7 we showed that for a two-player zero-sum game G, if both players use an expert algorithm with sublinear regret, then the average distribution converges to a Nash equilibrium. In this exercise you will prove that the same thing happens if only one player is using an expert algorithm, while the other is simply best responding. Formally, suppose the row player is using an expert algorithm to produce p_t such that

$$\sum_{t=1}^{T} G(p_t, q_t) \leq \min_p \sum_{t=1}^{T} G(p, q_t) + \mathcal{R}_T$$

for some sublinear regret $\mathcal{R}_T$, and the column player predicts $q_t = \operatorname{argmax}_q G(p_t, q)$. Then prove

$$\max_q G(\bar{p}, q) \leq v(G) + \frac{\mathcal{R}_T}{T} \quad \text{and} \quad \min_p G(p, \bar{q}) \geq v(G) - \frac{\mathcal{R}_T}{T}$$

where $\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p_t$, $\bar{q} = \frac{1}{T}\sum_{t=1}^{T} q_t$ and $v(G)$ is the value of the game. (This means that $(\bar{p}, \bar{q})$ converges to a Nash equilibrium.)

2. (**Smooth Games**) Consider a two-player non-zero-sum game with loss matrices $G_1 \in [0,1]^{N \times M}$ and $G_2 \in [0,1]^{N \times M}$, so that if player 1 plays action $i \in [N]$ and player 2 plays action $j \in [M]$, the losses for these two players are $G_1(i,j)$ and $G_2(i,j)$ respectively. The (negative) welfare for a profile (i,j) is simply the sum of the losses $G(i,j) \stackrel{\text{def}}{=} G_1(i,j) + G_2(i,j)$, and

$$G^\star = \min_{(i,j) \in [N] \times [M]} G_1(i,j) + G_2(i,j)$$

is the optimal welfare. The game is called (λ, μ)-smooth (for $\lambda > 0$ and $0 < \mu < 1$) if there exists a profile $(i^\star, j^\star)$ such that for any $(i,j) \in [N] \times [M]$,

$$G_1(i^\star, j) + G_2(i, j^\star) \leq \lambda G^\star + \mu G(i,j).$$

(a) Suppose both players use an expert algorithm to play a (λ, μ)-smooth game repeatedly with distribution $p_t \in \Delta(N)$ and $q_t \in \Delta(M)$ for time t, such that

$$\sum_{t=1}^{T} G_1(p_t, q_t) \leq \min_p \sum_{t=1}^{T} G_1(p, q_t) + \mathcal{R}_T \quad \text{and} \quad \sum_{t=1}^{T} G_2(p_t, q_t) \leq \min_q \sum_{t=1}^{T} G_2(p_t, q) + \mathcal{R}'_T.$$

for some sublinear regret terms $\mathcal{R}_T$ and $\mathcal{R}'_T$ (recall notation $G(p,q)$ from Lecture 7). Prove

$$\frac{1}{T}\sum_{t=1}^{T} G(p_t, q_t) \leq \frac{\lambda}{1-\mu} G^\star + \frac{\mathcal{R}_T + \mathcal{R}'_T}{(1-\mu)T},$$

that is, the average welfare converges to an approximate optimal value when $T \to \infty$.

(b) Suppose the expert algorithms used by the players actually enjoy "small-loss" bounds so that for any $i \in [N]$ and $j \in [M]$,

$$\sum_{t=1}^{T} G_1(p_t, q_t) \leq \sum_{t=1}^{T} G_1(i, q_t) + C\left(\sqrt{(\ln N)\sum_{t=1}^{T} G_1(i, q_t)} + \ln N\right)$$

and

$$\sum_{t=1}^{T} G_2(p_t, q_t) \le \sum_{t=1}^{T} G_2(p_t, j) + C\left(\sqrt{(\ln M)\sum_{t=1}^{T} G_2(p_t, j)} + \ln M\right).$$

for some absolute constant C. Prove that this allows the following faster convergence:

$$\frac{1}{T}\sum_{t=1}^{T} G(p_t, q_t) \le \frac{\lambda(1+\mu)}{\mu(1-\mu)} G^\star + \frac{C' \ln(NM)}{T},$$

for some constant C' depending on C and μ. (Hint, at some point you will need to solve a quadratic $ax^2 + bx + c \le 0$ with formula $x \le \frac{1}{2a}(-b + \sqrt{b^2 - 4ac}) \le \frac{1}{a}\sqrt{b^2 - 2ac}$. Try to identify what a, b, c and x are.)

3. Recall the AdaBoost algorithm discussed in Lecture 8. Prove that the distribution p_{t+1} is such that the previous classifier h_t has no edge over random guessing at all, that is,

$$\mathbb{E}_{i \sim p_{t+1}}\left[\mathbf{1}\{h_t(x_i) = y_i\}\right] = 1/2.$$

4. (**Weakly Adaptive Algorithm**) An OCO algorithm is called weakly adaptive if for any time interval $[s, e]$, we have (ignoring dependence on other parameters)

$$\sum_{t=s}^{e} f_t(w_t) - \min_{w \in \Omega} \sum_{t=s}^{e} f_t(w) = \mathcal{O}(\sqrt{T}).$$

Below you need to analyze several weakly adaptive algorithms.

(a) Prove that the following version of OGD (derived from the mirror descent framework from Homework 1):

$$w_{t+1} = \min_{w \in \Omega} \left\|w - w'_{t+1}\right\|_2^2 \quad \text{where} \quad w'_{t+1} = w_t - \eta \nabla f_t(w_t)$$

is already a weakly adaptive algorithm (!) with an appropriate choice of η. (Hint: reuse the analysis from Homework 1.)

(b) Since Hedge is also an instance of mirror descent, is Hedge also a weakly adaptive algorithm? Prove it if you think the answer is yes, otherwise construct a counter-example to disprove it and also point out why a proof similar to the last question does not work any more.

(c) In Lecture 11 we analyze the fixed-share algorithm by treating it as an instance of Hedge over a complicated space. Another way to analyze it is through the mirror descent framework. To see this, first write fixed-share as

$$p_t = (1-\alpha)\widetilde{p}_t + \frac{\alpha}{N}\mathbf{1}$$
$$\widetilde{p}_{t+1}(i) \propto p_t(i)\exp(-\eta \ell_t(i)) \quad \forall i,$$

By Question 6 (b) (d) from last homework, we know that the second update rule ensures

$$\langle p_t - q, \ell_t \rangle \le \frac{\mathrm{KL}(q, p_t) - \mathrm{KL}(q, \widetilde{p}_{t+1})}{\eta} + \eta. \tag{1}$$

Continue the proof to show that with $\alpha = 1/T$ and $\eta = \sqrt{\frac{\ln(NT)}{T}}$ we have

$$\sum_{t=s}^{e} \langle p_t - q, \ell_t \rangle = \mathcal{O}(\sqrt{T \ln(NT)}),$$

which means fixed-share is weakly adaptive.

(d) In last homework we also discuss the regret matching algorithm

$$p_{t+1}(i) \propto [R_t(i)]_+ \quad \text{where} \quad R_t(i) = R_{t-1}(i) + r_t(i)$$

(recall $r_t(i) = \langle p_t, \ell_t \rangle - \ell_t(i)$ and $[x]_+ = \max\{x, 0\}$). A simple upgrade of the algorithm, called regret-matching$^+$, predicts

$$p_{t+1}(i) \propto \widetilde{R}_t(i) \quad \text{where} \quad \widetilde{R}_t(i) = [\widetilde{R}_{t-1}(i) + r_t(i)]_+$$

which performs some sort of recursive clipping on the regret. Prove that for any $i \in [N]$,

$$\sum_{t=s}^{e} (\langle p_t, \ell_t \rangle - \ell_t(i)) \le \widetilde{R}_e(i) \le \sqrt{eN},$$

which implies that regret-matching$^+$ is also weakly adaptive (since $e \le T$).

5. (**Long-term Memory**) The switching regret we discussed in Lecture 10 only cares about the number of switches. What it does not capture is the possibly periodic phenomenon in practice. Consider the product recommendation example again where it is reasonable to imagine that a switch happens every month, but it makes sense to compete with the same policy for the same month of different years, meaning that the switches only happen between 12 policies . Intuitively, an algorithm with some sort of "long-term memory" should be able to exploit this periodic phenomenon and provide some more meaningful guarantee than the switching regret.

Formally, consider the expert problem and let $q_1, \dots, q_T \in \Delta(N)$ be a sequence of competitors such that $\sum_{t=2}^{T} \mathbf{1}\{q_t \neq q_{t-1}\} = S - 1$, but in addition the set $U = \{q_1, \dots, q_T\}$ has only n distinct elements for some $n \ll S$, which means there are many "switching-backs" happening. The regret is defined as usual

$$\mathcal{R}_T(q_1, \dots, q_T) = \sum_{t=1}^{T} \langle p_t - q_t, \ell_t \rangle .$$

(a) To first get a sense of how large the regret should be, consider the case when q_t only concentrates on one expert i_t for each t. Similar to the discussion of Lecture 11, we can create a set of meta-experts

$$\mathcal{M} = \{e \in [N]^T : \sum_{t=2}^{T} \mathbf{1}\{e(t) \neq e(t-1)\} = S - 1 \text{ and } |\{e(1), \dots, e(T)\}| = n\}$$

and run an expert algorithm with it. This will show that $\mathcal{R}_T(i_1, \dots, i_T) \le \mathcal{O}(\sqrt{T \ln |\mathcal{M}|})$. Write down the concrete regret bound by figuring out how large $|\mathcal{M}|$ is (recall $\ln \binom{b}{a} = \mathcal{O}(a \ln \frac{b}{a})$).

(b) To get roughly the same bound efficiently, we generalize the fixed-share algorithm as:

$$p_t = \sum_{\tau=1}^{t} \alpha_t(\tau) \widetilde{p}_\tau$$
$$\widetilde{p}_{t+1}(i) \propto p_t(i) \exp(-\eta \ell_t(i)) \quad \forall i$$

with $\widetilde{p}_1$ being the uniform distribution and $\alpha_t \in \Delta(t)$ be some distribution over the history. In other words, generalized fixed-share is mixing the past predictions to obtain long-term memory. The regular fixed-share is clearly is special case when $\alpha_t(t) = 1 - \alpha$ and $\alpha_t(1) = \alpha$ for some α, that is, only mixing the current distribution with the initial uniform distribution.

(i) Let $s_t = \max\{s \in [T] : s < t, q_s = q_t\}$ be the most recent past appearance time of competitor q_t (if such time does not exist, that is, the set in the definition is empty, then s_t is defined as 0). Use Eq. (1) (which clearly still holds) to show

$$\langle p_t - q_t, \ell_t \rangle \le \frac{\ln\left(\frac{1}{\alpha_t(s_t+1)}\right) + \mathrm{KL}(q_t, \widetilde{p}_{s_t+1}) - \mathrm{KL}(q_t, \widetilde{p}_{t+1})}{\eta} + \eta. \tag{2}$$

(ii) Conclude the following regret bound by summing up Eq. (2):

$$\mathcal{R}_T(q_1, \dots, q_T) \le \frac{1}{\eta} \sum_{t=1}^{T} \ln\left(\frac{1}{\alpha_t(s_t+1)}\right) + \frac{n \ln N}{\eta} + \eta T.$$

(iii) (**Uniform Past Mixing**) For $t \geq 2$, let $\alpha_t(t) = 1 - \alpha$ and $\alpha_t(\tau) = \alpha/(t-1),\ \forall \tau < t$ for some parameter α. Show that with the optimal tunning of α and η, generalized fixed-share ensures

$$\mathcal{R}_T(q_1, \ldots, q_T) = \mathcal{O}\left(\sqrt{T(S \ln T + n \ln N)}\right).$$

(Hint: recall the fact $a \ln \frac{1}{x} + b \ln \frac{1}{1-x}$ for $x \in (0,1)$ is minimized when $x = \frac{a}{a+b}$ and the optimal value is $\mathcal{O}(a \ln \frac{a+b}{a})$ when $a < b$.)

(iv) (**Decaying Past Mixing**) For $t \geq 2$, let $\alpha_t(t) = 1 - \alpha$ and $\alpha_t(\tau) = \frac{\alpha}{(t-\tau)Z_t},\ \forall \tau < t$ for some parameter α and normalization factor $Z_t = \sum_{\tau=1}^{t-1} \frac{1}{t-\tau} = \mathcal{O}(\ln T)$. Show that with the optimal tunning of α and η, generalized fixed-share ensures

$$\mathcal{R}_T(q_1, \ldots, q_T) = \mathcal{O}\left(\sqrt{T\left(S \ln(\ln T) + S \ln\left(\frac{nT}{S}\right) + n \ln N\right)}\right).$$