CSCI 659 Homework 2

Spring 2026

Instructor: Haipeng Luo

This homework is due on **03/13, 11:59pm**. *See course website for more instructions on finishing and submitting your homework as well as the late policy. Total points: 60.*

1. (**Optimistic OMD**) Similar to the optimistic version of FTRL, the general Online Mirror Descent framework that we saw in HW1 can also incorporate an arbitrary loss predictor m_t at round t. The resulting algorithm, called Optimistic Online Mirror Descent, starts with an arbitrary $w'_1 \in \Omega$, and plays at time t the following action:

$$w_t = \operatorname*{argmin}_{w\in\Omega} \langle w, m_t\rangle + \tfrac{1}{\eta} D_\psi(w, w'_t),$$

followed by the an update to the auxiliary action w'_t after seeing the true loss vector ℓ_t:

$$w'_{t+1} = \operatorname*{argmin}_{w\in\Omega} \langle w, \ell_t\rangle + \tfrac{1}{\eta} D_\psi(w, w'_t),$$

where as before $\eta > 0$ is a learning rate and ψ is a convex regularizer. Note that the auxiliary sequence $w'_1, \ldots, w'_T$ is exactly the output of the vanilla OMD. Unlike Optimistic FTRL where such an auxiliary sequence only appears in the analysis, Optimistic OMD needs to explicitly maintain this sequence in order to compute the actual action sequence $w_1, \ldots, w_T$.

(a) (4pts) Based on Eq. (1) of HW1, we clearly have for any $u \in \Omega$:

$$\eta \left\langle w'_{t+1} - u, \ell_t\right\rangle \le D_\psi(u, w'_t) - D_\psi(u, w'_{t+1}) - D_\psi(w'_{t+1}, w'_t). \tag{1}$$

Follow the same proof of this fact to show that for any $u' \in \Omega$:

$$\eta \langle w_t - u', m_t\rangle \le D_\psi(u', w'_t) - D_\psi(u', w_t) - D_\psi(w_t, w'_t). \tag{2}$$

Then, combine these two inequalities and choose a specific u' to further prove for any $u \in \Omega$:

$$\begin{aligned} \langle w_t - u, \ell_t\rangle \le \left\langle w_t - w'_{t+1}, \ell_t - m_t\right\rangle &+ \tfrac{1}{\eta} D_\psi(u, w'_t) - \tfrac{1}{\eta} D_\psi(u, w'_{t+1}) \\ &- \tfrac{1}{\eta} D_\psi(w'_{t+1}, w_t) - \tfrac{1}{\eta} D_\psi(w_t, w'_t). \end{aligned} \tag{3}$$

(b) (6pts) Suppose that ψ is strongly convex with respect to some norm $\|\cdot\|$. Make use of Eq. (1) and Eq. (2) again to prove the following stability statement (which optimistic FTRL also enjoys):

$$\left\| w_t - w'_{t+1} \right\| \le \eta \left\| \ell_t - m_t \right\|_\star . \tag{4}$$

Combine everything to show the final regret bound:

$$\mathcal{R}_T \le \frac{\max_{u\in\Omega} D_\psi(u, w'_1)}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t - m_t\|_\star^2 - \frac{1}{4\eta} \sum_{t=2}^{T} \|w_t - w_{t-1}\|^2.$$

(Note the similarity of this bound compared to that of Theorem 1 from Lecture 3 for Optimistic FTRL.)

2. (**Learning in Zero-Sum Games**) Consider the learning setup for a two-player zero-sum game $G \in [0,1]^{N\times M}$ discussed in Lecture 4: at each round $t = 1, \ldots, T$, row player uses a no-regret algorithm to come up with $p_t \in \Delta(N)$ and the column player comes up with $q_t \in \Delta(M)$ in some way, after which the row player suffers loss $G(p_t, q_t)$ and see $G(i, q_t)$ for all i. Let the empirical average strategies be $\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p_t$ and $\bar{q} = \frac{1}{T}\sum_{t=1}^{T} q_t$.

   (a) (4pts) Suppose that the column player gets to see p_t before deciding q_t. Then naturally she would choose to best respond to p_t, that is, $q_t \in \operatorname{argmax}_{q\in\Delta(M)} G(p_t, q)$. In this case, prove that $(\bar{p}, \bar{q})$ is an ϵ-approximate Nash equilibrium with $\epsilon = \frac{\mathcal{R}_T}{T}$ (where $\mathcal{R}_T = \sum_{t=1}^{T} G(p_t, q_t) - \min_p \sum_{t=1}^{T} G(p, q_t)$ is the regret of the row player), that is, $\max_q G(\bar{p}, q) - \epsilon \leq G(\bar{p}, \bar{q}) \leq \min_p G(p, \bar{q}) + \epsilon$.

   (b) (5pts) Now suppose that both the row player and the column player use a no-regret algorithm such that, no matter what their opponent plays, their regret is bounded by $B(T)$ and $B'(T)$ respectively for some function B and B' (e.g. $B(T) = \sqrt{T \ln N}$ and $B'(T) = \sqrt{T \ln M}$). In Lecture 1, we mentioned that the definition of (static) regret might not be very reasonable when against an adaptive adversary whose decisions would change accordingly if the learner were to stick with a fixed action for all rounds. In the literature, a different regret measure, called *policy regret*, was exactly proposed to address this issue. Specifically, in this game setting, the policy regret $\mathcal{PR}_T$ of the row player is defined as

$$\mathcal{PR}_T = \sum_{t=1}^{T} G(p_t, q_t) - \min_{p\in\Delta(N)} \sum_{t=1}^{T} G(p, q_t^{(p)})$$

   where $q_t^{(p)}$ is what the column player would have played at time t if the row player were to play p all the time. Show that $\mathcal{PR}_T \leq B(T) + B'(T)$ (and thus, given that the same should also hold for the column player by symmetry, if both of them try to selfishly minimize their policy regret, they might just as well minimize their own static regret).

3. (**Weakly Adaptive Algorithm**) Recall that an OCO algorithm is called weakly adaptive if for any time interval $\mathcal{I} = [s, e]$, its interval regret satisfies $\mathcal{R}_\mathcal{I} = \widetilde{\mathcal{O}}(\sqrt{T})$ (ignoring dependence on other parameters). Below, you need to analyze several weakly adaptive algorithms.

   (a) In HW1 we discussed the Regret Matching (RM) algorithm for the expert problem:
$$p_{t+1}(i) \propto [R_t(i)]_+$$
   where $R_t(i) = R_{t-1}(i) + r_t(i)$ (with $R_0 = \mathbf{0}$), $r_t(i) = \langle p_t, \ell_t \rangle - \ell_t(i)$, and $[x]_+ = \max\{x, 0\}$. A simple upgrade of the algorithm, called Regret Matching+ (RM+), makes the following modification:
$$p_{t+1}(i) \propto \widetilde{R}_t(i)$$
   where $\widetilde{R}_t(i) = [\widetilde{R}_{t-1}(i) + r_t(i)]_+$ (with $\widetilde{R}_0 = \mathbf{0}$). (See if you can spot the similarity between RM versus RM+ and lazy OGD versus non-lazy OGD.)

   (i) (4pts) Follow the same ideas from HW1 Problems 2(a) and 2(b) to show that for any time step e, we have $\sum_{i=1}^N \widetilde{R}_e^2(i) \le eN$.

   (ii) (4pts) Further prove $\widetilde{R}_e(i) \ge \sum_{t=s}^e r_t(i)$ for any expert i and any starting time step $s \le e$, and conclude that RM+ is a weakly adaptive algorithm.

   (b) Next, consider running OMD with a strongly convex regularizer (see Problem 1 with $m_t = \mathbf{0}$ so that $w_t = w_t'$). Combining Eq. (3) and Eq. (4) and dropping some nonpositive terms, we have the following per-round regret bound
$$\langle w_t - u, \ell_t \rangle \le \tfrac{1}{\eta} D_\psi(u, w_t) - \tfrac{1}{\eta} D_\psi(u, w_{t+1}) + \eta G^2.$$
   where G is such that $\|\ell_t\|_\star \le G$ for all t.

   (i) (3pts) Starting from this per-round regret bound, prove that for any interval $\mathcal{I} = [s, e]$, we have the following interval regret bound
$$\mathcal{R}_\mathcal{I} = \max_{u \in \Omega} \sum_{t=s}^{e} \langle w_t - u, \ell_t \rangle \le \frac{1}{\eta} \bar{B}_\psi + \eta |\mathcal{I}| G^2,$$
   where $\bar{B}_\psi = \max_{w, w' \in \Omega} D_\psi(w, w')$, and then pick an appropriate η independent of $\mathcal{I}$ to further conclude that OMD is weakly adaptive as long as $\bar{B}_\psi$ is finite. (You are encouraged to think about whether FTRL enjoys a similar result.)

   (ii) (6pts) From HW1 we know that OGD is OMD with $\psi(w) = \frac{1}{2}\|w\|_2^2$, and thus $\bar{B}_\psi = \max_{w,w' \in \Omega} \frac{1}{2}\|w - w'\|_2^2$ is bounded for any bounded decision set Ω. On the other hand, Hedge is OMD with $\psi(p) = \sum_{i=1}^N p(i) \ln p(i)$, and $\bar{B}_\psi$ becomes unbounded since $D_\psi(p, q) = \sum_{i=1}^N p(i) \ln \frac{p(i)}{q(i)}$ is the KL divergence between two distributions and can be unbounded if $q(i) = 0$ for some i in the support of p. One way to fix this is to change the decision set of OMD from the simplex $\Delta(N)$ to a clipped simplex $\Omega = \{p \in \Delta(N) : p(i) \ge \delta, \ \forall i\}$ for some parameter $\delta \in (0, 1/N)$. That is, the algorithm is now
$$p_{t+1} = \operatorname*{argmin}_{p \in \Omega} \langle p, \ell_t \rangle + \frac{1}{\eta} D_\psi(p, p_t) = \operatorname*{argmin}_{p \in \Omega} \langle p, \ell_t \rangle + \frac{1}{\eta} \sum_{i=1}^N p(i) \ln \frac{p(i)}{p_t(i)}.$$
   Pick an appropriate δ and an appropriate η (both independent of $\mathcal{I}$) to show that the algorithm above enjoys for all interval $\mathcal{I}$:
$$\mathcal{R}_\mathcal{I} = \max_{p \in \Delta(N)} \sum_{t=s}^{e} \langle p_t - p, \ell_t \rangle = \mathcal{O}(\sqrt{T \ln(NT)}).$$
   (Hint: make sure to consider the difference between $\mathcal{R}_\mathcal{I}$ and $\max_{p \in \Omega} \sum_{t=s}^e \langle p_t - p, \ell_t \rangle$.)

4. (**Confidence-Rated Experts**) The confidence-rated expert problem is a generalization of the sleeping expert problem. Instead of being either asleep ($a_t(i) = 0$) or awake ($a_t(i) = 1$) at each round t, each expert can provide a "confidence score" $a_t(i) \in [0, 1]$ for the advice that she provides for this round (the larger the score, the more confident the expert). Formally, the learning protocol is as follows: for each round $t = 1, \ldots, T$,

   - each expert i provides an arbitrary confidence score $a_t(i) \in [0, 1]$, revealed to the learner;
   - the learner decides a distribution $p_t \in \Delta(N)$ with the restriction that no weights are put on zero-confidence experts, that is, $p_t(i) = 0$ if $a_t(i) = 0$;
   - the environment decides and reveals the loss $\ell_t(i)$ for each expert i.[1]

   The confidence-rated regret of the learner against expert i is defined as $R_T(i) = \sum_{t=1}^{T} a_t(i) \langle p_t - e_i, \ell_t \rangle$ (where e_i is the i-th basis vector). This is clearly a direct generalization of the sleeping expert problem from binary $a_t(i)$ to real value $a_t(i)$.

   In fact, one can also directly generalize the reduction from sleeping experts to regular experts discussed in Lecture 5 to this case. The resulting algorithm is as follows (convince yourself that this indeed recovers Algorithm 1 of Lecture 5 when $a_t(i)$ is binary).

Algorithm 1: Reduction from Confidence-Rate Experts to Regular Experts

```
Input: a regular expert algorithm $\mathcal{E}$
for $t = 1, \ldots, T$ do
    let $\widehat{p}_t \in \Delta(N)$ be the decision of $\mathcal{E}$ at round $t$
    observe $a_t$ from the environment
    play $p_t \in \Delta(N)$ such that $p_t(i) \propto a_t(i)\widehat{p}_t(i)$
    observe the loss vector $\ell_t$
    set $\widehat{\ell}_t(i) = a_t(i)\ell_t(i) + (1 - a_t(i)) \langle p_t, \ell_t \rangle$ for all $i$
    feed $\widehat{\ell}_t$ to $\mathcal{E}$
```

(a) (4pts) Prove that for each t, $\left\langle \widehat{p}_t, \widehat{\ell}_t \right\rangle = \langle p_t, \ell_t \rangle$ holds.

(b) (3pts) Further show that for each t and i, $a_t(i) \langle p_t - e_i, \ell_t \rangle = \left\langle \widehat{p}_t - e_i, \widehat{\ell}_t \right\rangle$ holds.

(c) (2pts) Finally, suppose that the given regular expert algorithm $\mathcal{E}$ ensures for each i: $\sum_{t=1}^{T} \left\langle \widehat{p}_t - e_i, \widehat{\ell}_t \right\rangle = 2\sqrt{\sum_{t=1}^{T} \left\langle \widehat{p}_t - e_i, \widehat{\ell}_t \right\rangle^2 \ln N}$ (basically the same as Eq. (1) of Lecture 5). Prove that Algorithm 1 ensures for each i:

$$R_T(i) = \sum_{t=1}^{T} a_t(i) \langle p_t - e_i, \ell_t \rangle = 2\sqrt{\sum_{t=1}^{T} a_t^2(i) \langle p_t - e_i, \ell_t \rangle^2 \ln N}. \tag{5}$$

[1] For simplicity, we assume that the loss is defined for zero-confidence experts as well. One can verify that this does not affect any of the following discussions.

5. (**Long-Term Memory**) The switching regret bound $\widetilde{\mathcal{O}}(\sqrt{ST})$ discussed in Lecture 5 only cares about the number of switches S. What it does not capture is the possibly periodic phenomenon in practice. For example, in a recommendation system, one can imagine that a switch happens in users' preferences every now and then (e.g. every season), but it is reasonable to imagine that the current preferences are similar to some preferences in the past (e.g. this spring is similar to the last spring). Intuitively, an algorithm with some kind of "long-term memory" should be able to exploit this periodic phenomenon and provide a better guarantee.

   Formally, consider the expert problem and let $j_1, \ldots, j_T \in [N]$ be a sequence of comparators such that there are $S-1$ switches: $\sum_{t=2}^{T} \mathbf{1}\{j_t \neq j_{t-1}\} = S - 1$, but in addition the set $U = \{j_1, \ldots, j_T\}$ only has n distinct elements for some $n \ll S$, implying that there are many "switching-backs" in this sequence. In this exercise, you need to analyze an algorithm whose dynamic regret against such a sequence $\mathcal{R}_T(j_1, \ldots, j_T) = \sum_{t=1}^{T} \langle p_t - e_{j_t}, \ell_t \rangle$ is of order $\mathcal{O}(\sqrt{T(S\ln T + n\ln N)})$, improving the typical switching regret bound $\mathcal{O}(\sqrt{TS\ln(NT)}) = \mathcal{O}(\sqrt{T(S\ln T + S\ln N)})$.

   The algorithm is again by reduction. In particular, we make use of a confidence-rated expert algorithm $\mathcal{C}$ (see Problem 4) working over the N experts, and N regular expert algorithms $\mathcal{E}_1, \ldots, \mathcal{E}_N$, each of which working on two imaginary experts called "Awake" and "Asleep". At each round t, each $\mathcal{E}_i$ proposes a distribution over these two imaginary experts, which we denote as $q_t^i = (a_t(i), 1 - a_t(i)) \in \Delta(2)$. The complete reduction is shown below.

Algorithm 2: Reduction for Long-Term Memory

```
Input: a parameter $\eta \in (0, 1/5]$, a confidence-rated expert algorithm $\mathcal{C}$, and $N$ regular expert
algorithms $\mathcal{E}_1, \ldots, \mathcal{E}_N$ working over the two imaginary experts "Awake" and "Asleep"
for $t = 1, \ldots, T$ do
    $\forall i$, obtain distribution $q_t^i = (a_t(i), 1 - a_t(i)) \in \Delta(2)$ from $\mathcal{E}_i$
    treat $a_t(1), \ldots, a_t(N)$ as the confidence scores and feed them to $\mathcal{C}$
    obtain and play the distribution $p_t \in \Delta(N)$ output from $\mathcal{C}$
    observe the loss vector $\ell_t \in [0, 1]^N$
    feed $\ell_t$ to $\mathcal{C}$
    $\forall i$, feed the loss vector $g_t^i = (5\eta - r_t(i), 0) \in [-1, 2]^2$ to $\mathcal{E}_i$ where $r_t(i) = \langle p_t - e_i, \ell_t \rangle$
```

   (a) (3pts) Suppose that the confidence-rated expert algorithm $\mathcal{C}$ satisfies Eq. (5). Prove that for each $j \in U = \{j_1, \ldots, j_T\}$, we have

$$\sum_{t=1}^{T} a_t(j) r_t(j) \leq 2\sqrt{\sum_{t=1}^{T} a_t(j) \ln N} \leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} a_t(j).$$

   (b) (6pts) Suppose that each regular expert algorithm $\mathcal{E}_i$ ensures the following switching regret bound: for any sequence $b_1, \ldots, b_T \in \{1, 2\}$ with $1 + \sum_{t=2}^{T} \mathbf{1}\{b_t \neq b_{t-1}\} = S_b$,

$$\sum_{t=1}^{T} \langle q_t^i, g_t^i \rangle - \sum_{t=1}^{T} g_t^i(b_t) \leq \frac{S_b \ln T}{\eta} + \eta \sum_{t=1}^{T} \sum_{k \in \{1,2\}} q_t^i(k) g_t^i(k)^2$$

   (you are encouraged to think about which algorithms actually satisfy this). Then, for each $j \in U$, pick an appropriate sequence of $b_1, \ldots, b_T$, apply this switching regret bound, and rearrang to show

$$\sum_{t : j_t = j} r_t(j) \leq \sum_{t=1}^{T} a_t(j) r_t(j) - \eta \sum_{t=1}^{T} a_t(j) + \frac{S_j \ln T}{\eta} + 5\eta T_j$$

   where $T_j = |\{t : j_t = j\}|$ and $S_j = 1 + \sum_{t=2}^{T} \mathbf{1}\{\text{exactly one of } j_t \text{ and } j_{t-1} \text{ is } j\}$.

(c) (6pts) Combine the results from the last two questions to conclude the dynamic regret bound

$$\mathcal{R}_T(j_1, \dots, j_T) = \sum_{t=1}^{T} \langle p_t - e_{j_t}, \ell_t \rangle = \mathcal{O}(\sqrt{T(S \ln T + n \ln N)})$$

when η is optimally tuned (which can depend on everything including S and n).