CSCI 659 Homework 2

Fall 2022

Instructor: Haipeng Luo

This homework is due on **10/16, 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 $u \in \Omega$:

$$\eta \langle w_{t+1}' - u, \ell_t \rangle \leq D_\psi(u, w_t') - D_\psi(u, w_{t+1}') - D_\psi(w_{t+1}', w_t'). \tag{1}$$

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

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

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

$$\begin{aligned} \langle w_t - u, \ell_t \rangle \leq {} & \langle w_t - w_{t+1}', \ell_t - m_t \rangle + \tfrac{1}{\eta} D_\psi(u, w_t') - \tfrac{1}{\eta} D_\psi(u, w_{t+1}') \\ & - \tfrac{1}{\eta} D_\psi(w_{t+1}', w_t) - \tfrac{1}{\eta} D_\psi(w_t, w_t'). \end{aligned} \tag{3}$$

Proof. Let $F(w) = \langle w, m_t \rangle + \tfrac{1}{\eta} D_\psi(w, w_t')$. Since w_t minimizes F, by applying Lemma 1 of Lecture 2 we have for any $u' \in \Omega$:

$$\langle w_t, m_t \rangle + \frac{1}{\eta} D_\psi(w_t, w_t') \leq \langle u', m_t \rangle + \frac{1}{\eta} D_\psi(u', w_t') - D_F(u', w_t).$$

Using $D_F = \tfrac{1}{\eta} D_\psi$ and rearranging then proves Eq. (2). Eq. (3) can be obtained by choosing $u' = w_{t+1}'$, adding up the two inequalities, adding $\langle w_t, \ell_t \rangle$ to both sides, and rearranging. $\square$

(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):

$$\|w_t - w_{t+1}'\| \leq \eta \|\ell_t - m_t\|_\star. \tag{4}$$

Combine everything to show the final regret bound:

$$\mathcal{R}_T \leq \frac{\max_{u \in \Omega} D_\psi(u, w_1')}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t - m_t\|_\star^2 - \frac{1}{4\eta} \sum_{t=2}^{T} \|w_t - w_{t-1}\|^2.$$

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

Proof. Summing up Eq. (1) and Eq. (2) with $u = w_t$ and $u' = w_{t+1}'$, we have

$$\eta \langle w_{t+1}' - w_t, \ell_t - m_t \rangle \leq -D_\psi(w_t, w_{t+1}') - D_\psi(w_{t+1}', w_t) \leq -\left\|w_t - w_{t+1}'\right\|^2,$$

where the last step is by the strong convexity of ψ. Rearranging and using Hölder's inequality, we arrive at

$$\left\|w_t - w_{t+1}'\right\|^2 \leq \eta \langle w_t - w_{t+1}', \ell_t - m_t \rangle \leq \eta \left\|w_t - w_{t+1}'\right\| \|\ell_t - m_t\|_\star.$$

Dividing both sides by $\left\|w_t - w_{t+1}'\right\|$ proves Eq. (4). Combining this with Eq. (3), using Hölder's inequality again, summing over t, and telescoping shows

$$\mathcal{R}_T \leq \frac{\max_{u \in \Omega} D_\psi(u, w_1')}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t - m_t\|_\star^2 - \frac{1}{\eta} D_\psi(w_{t+1}', w_t) - \frac{1}{\eta} D_\psi(w_t, w_t').$$

The last step of lower bounding $\sum_{t=1}^{T} D_\psi(w_{t+1}', w_t) + D_\psi(w_t, w_t')$ by $\frac{1}{4} \sum_{t=2}^{T} \|w_t - w_{t-1}\|^2$ is exactly the same as the proof for Theorem 1 of Lecture 3. $\square$

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, $q_t \in \mathrm{argmax}_{q \in \Delta(M)} G(p_t, q)$. In this case, prove that $(\bar{p}, \bar{q})$ is an approximate Nash equilibrium with error $\epsilon = \frac{\mathcal{R}_T}{T}$ (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, $\max_q G(\bar{p}, q) - \epsilon \le G(\bar{p}, \bar{q}) \le \min_p G(p, \bar{q}) + \epsilon$.

Proof. In Lecture 4, we know that $(\bar{p}, \bar{q})$ is an approximate Nash equilibrium with error $\epsilon = \frac{\mathcal{R}_T + \mathcal{R}'_T}{T}$ where $\mathcal{R}'_T$ is the regret of the column player. Thus, it suffices to show that the column player has no positive regret at all, which is indeed the case by simply using the definition of q_t:

$$\mathcal{R}'_T = \max_q \sum_{t=1}^{T} G(p_t, q) - \sum_{t=1}^{T} G(p_t, q_t) = \max_q \sum_{t=1}^{T} G(p_t, q) - \sum_{t=1}^{T} \max_q G(p_t, q) \le 0.$$

$\square$

(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 $\mathcal{PR}_T \le B(T) + B'(T)$ (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).

Proof. For any $p \in \Delta(N)$, when the row player always plays p, the column player ensures the following static regret bound

$$\max_{q \in \Delta(M)} \sum_{t=1}^{T} G(p, q) - \sum_{t=1}^{T} G(p, q_t^{(p)}) \le B'(T).$$

Therefore, the policy regret of the row player can be bounded as

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

$$\le \sum_{t=1}^{T} G(p_t, q_t) - \min_{p \in \Delta(N)} \max_{q \in \Delta(M)} \sum_{t=1}^{T} G(p, q) + B'(T)$$

$$\le \sum_{t=1}^{T} G(p_t, q_t) - \min_{p \in \Delta(N)} \sum_{t=1}^{T} G(p, \bar{q}) + B'(T)$$

$$= \sum_{t=1}^{T} G(p_t, q_t) - \min_{p \in \Delta(N)} \sum_{t=1}^{T} G(p, q_t) + B'(T)$$

$$\le B(T) + B'(T),$$

which completes the proof. $\square$

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 $\sum_{i=1}^N \widetilde{R}_e^2(i) \le eN$.

Proof. By the exact same reasoning of HW1 Problem 2(a), we first have

$$\widetilde{R}_t^2(i) \le \widetilde{R}_{t-1}^2(i) + 2\widetilde{R}_{t-1}(i)r_t(i) + r_t^2(i).$$

Then, by the same reasoning of Problem 2(b), we have $\sum_{i=1}^N \widetilde{R}_{t-1}(i)r_t(i) = 0$ by the fact $p_t(i) \propto \widetilde{R}_{t-1}(i)$ and

$$\sum_{i=1}^N p_t(i)r_t(i) = \sum_{i=1}^N p_t(i)\langle p_t, \ell_t \rangle - \sum_{i=1}^N p_t(i)\ell_t(i) = \langle p_t, \ell_t \rangle - \langle p_t, \ell_t \rangle = 0.$$

Thus, we conclude with $\sum_{i=1}^N \widetilde{R}_e^2(i) \le \sum_{i=1}^N \widetilde{R}_{e-1}^2(i) + N \le \cdots \le eN$. $\square$

(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.

Proof. By definition we have $\widetilde{R}_t(i) \ge \widetilde{R}_{t-1}(i) + r_t(i)$ and $\widetilde{R}_t(i) \ge 0$ for any t, which implies

$$\widetilde{R}_e(i) \ge \widetilde{R}_{e-1}(i) + r_e(i) \ge \cdots \ge \widetilde{R}_{s-1}(i) + \sum_{t=s}^e r_t(i) \ge \sum_{t=s}^e r_t(i).$$

Therefore, the interval regret of RM+ satisfies

$$\mathcal{R}_\mathcal{I} = \max_{i\in[N]} \sum_{t=s}^e r_t(i) \le \max_{i\in[N]} \widetilde{R}_e(i) \le \sqrt{\sum_{i=1}^N \widetilde{R}_e^2(i)} \le \sqrt{eN} \le \sqrt{TN},$$

showing that RM+ is indeed a weakly adaptive algorithm. $\square$

(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.)

4

Proof. Simply summing over t from s to e and telescoping shows

$$\mathcal{R}_\mathcal{I} \leq \frac{1}{\eta} D_\psi(u, w_s) - \frac{1}{\eta} D_\psi(u, w_{e+1}) + \eta |\mathcal{I}| G^2 \leq \frac{1}{\eta} D_\psi(u, w_s) + \eta |\mathcal{I}| G^2.$$

Further bounding $D_\psi(u, w_s)$ by $\bar{B}_\psi$ and picking $\eta = \sqrt{\frac{\bar{B}_\psi}{TG^2}}$ proves $\mathcal{R}_\mathcal{I} = \mathcal{O}(G\sqrt{\bar{B}_\psi T})$, that is, the algorithm is weakly adaptive. $\qquad\square$

(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 $\Omega = \{p \in \Delta(N) : p(i) \geq \delta, \ \forall i\}$ for some parameter $\delta \in (0, 1/N)$. That is, the algorithm is now

$$p_{t+1} = \underset{p \in \Omega}{\operatorname{argmin}} \langle p, \ell_t \rangle + \frac{1}{\eta} D_\psi(p, p_t) = \underset{p \in \Omega}{\operatorname{argmin}} \langle p, \ell_t \rangle + \frac{1}{\eta} \sum_{i=1}^N p(i) \ln \frac{p(i)}{p_t(i)}.$$

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 $\max_{p \in \Omega} \sum_{t=s}^e \langle p_t - p, \ell_t \rangle$.)

Proof. With the new domain, $\bar{B}_\psi$ is now bounded as

$$\bar{B}_\psi = \max_{p,q \in \Omega} \sum_{i=1}^N p(i) \ln \frac{p(i)}{q(i)} \leq \max_{p \in \Omega} \sum_{i=1}^N p(i) \ln \frac{p(i)}{\delta} \leq \ln \frac{1}{\delta}.$$

Therefore, based on Problem (i), we have (recall $G = 1$ in this case)

$$\max_{p \in \Omega} \sum_{t=s}^e \langle p_t - p, \ell_t \rangle \leq \frac{1}{\eta} \ln \frac{1}{\delta} + \eta T.$$

Next, we consider the difference between $\mathcal{R}_\mathcal{I}$ and $\max_{p \in \Omega} \sum_{t=s}^e \langle p_t - p, \ell_t \rangle$. Let $p^\star \in \min_{p \in \Delta(N)} \sum_{t=s}^e \langle p, \ell_t \rangle$ and $p_\delta^\star = (1 - N\delta)p^\star + \delta\mathbf{1} \in \Omega$ where $\mathbf{1}$ is the all-one vector. Then we bound $\mathcal{R}_\mathcal{I}$ as

$$\begin{aligned}
\mathcal{R}_\mathcal{I} &= \sum_{t=s}^e \langle p_t - p^\star, \ell_t \rangle = \sum_{t=s}^e \left\langle p_t - \frac{1}{1 - N\delta}(p_\delta^\star - \delta\mathbf{1}), \ell_t \right\rangle \\
&\leq \sum_{t=s}^e \langle p_t - (p_\delta^\star - \delta\mathbf{1}), \ell_t \rangle \leq \max_{p \in \Omega} \sum_{t=s}^e \langle p_t - p, \ell_t \rangle + \delta \sum_{t=s}^e \langle \mathbf{1}, \ell_t \rangle \\
&\leq \frac{1}{\eta} \ln \frac{1}{\delta} + \eta T + \delta N T.
\end{aligned}$$

Therefore, by picking $\delta = \frac{1}{NT}$ and $\eta = \sqrt{\frac{\ln(NT)}{T}}$ (or any other suitable choices), we have shown $\mathcal{R}_\mathcal{I} = \mathcal{O}(\sqrt{T \ln(NT)})$. $\qquad\square$

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 ℓ_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.

Proof. We proceed using definitions:

$$\left\langle \widehat{p}_t, \widehat{\ell}_t \right\rangle = \sum_{i=1}^{N} \widehat{p}_t(i)a_t(i)\ell_t(i) + \sum_{i=1}^{N} \widehat{p}_t(i)(1 - a_t(i)) \langle p_t, \ell_t \rangle$$

$$= \langle p_t, \ell_t \rangle + \sum_{i=1}^{N} \widehat{p}_t(i)a_t(i)(\ell_t(i) - \langle p_t, \ell_t \rangle) = \langle p_t, \ell_t \rangle,$$

where the last step is by the facts $p_t(i) \propto a_t(i)\widehat{p}_t(i)$ and $\sum_{i=1}^{N} p_t(i)(\ell_t(i) - \langle p_t, \ell_t \rangle) = 0$ (the latter of which has been used in analyzing RM/RM+). $\square$

(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.

Proof. Again, using definitions and the result from the last question, we have

$$a_t(i) \langle p_t - e_i, \ell_t \rangle = a_t(i) \langle p_t, \ell_t \rangle - a_t(i)\ell_t(i)$$

$$= a_t(i) \langle p_t, \ell_t \rangle - \widehat{\ell}_t(i) + (1 - a_t(i)) \langle p_t, \ell_t \rangle$$

$$= \langle p_t, \ell_t \rangle - \widehat{\ell}_t(i) = \left\langle \widehat{p}_t, \widehat{\ell}_t \right\rangle - \widehat{\ell}_t(i) = \left\langle \widehat{p}_t - e_i, \widehat{\ell}_t \right\rangle.$$

$\square$

---

[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.

(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:

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

Proof. This is simply by applying the result from the last question twice:

$$R_T(i) = \sum_{t=1}^{T} \left\langle \widehat{p}_t - e_i, \widehat{\ell}_t \right\rangle = 2\sqrt{\sum_{t=1}^{T} \left\langle \widehat{p}_t - e_i, \widehat{\ell}_t \right\rangle^2 \ln N}$$

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

$\square$

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**

$\quad \forall i$, obtain distribution $q_t^i = (a_t(i), 1 - a_t(i)) \in \Delta(2)$ from $\mathcal{E}_i$

$\quad$ treat $a_t(1), \ldots, a_t(N)$ as the confidence scores and feed them to $\mathcal{C}$

$\quad$ obtain and play the distribution $p_t \in \Delta(N)$ output from $\mathcal{C}$

$\quad$ observe the loss vector $\ell_t \in [0, 1]^N$

$\quad$ feed ℓ_t to $\mathcal{C}$

$\quad \forall i$, 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

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

Proof. This is a direct consequence of Eq. (5):

$$\sum_{t=1}^{T} a_t(j) r_t(j) = \sum_{t=1}^{T} a_t(j) \langle p_t - e_j, \ell_t \rangle$$

$$\leq 2\sqrt{\sum_{t=1}^{T} a_t^2(j) \langle p_t - e_j, \ell_t \rangle^2 \ln N}$$

$$\leq 2\sqrt{\sum_{t=1}^{T} a_t(j) \ln N} \qquad (a_t(j) \in [0, 1] \text{ and } |\langle p_t - e_j, \ell_t \rangle| \leq 1)$$

$$\leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} a_t(j). \qquad (2xy \leq x^2 + y^2)$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

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

Proof. Let $b_t = 1$ if $j_t = j$ and $b_t = 2$ otherwise. Then S_b is exactly S_j, and with the definition of q_t^i and g_t^i, the switching regret bound implies:

$$\sum_{t=1}^{T} a_t(j)(5\eta - r_t(j)) - \sum_{t:j_t=j} (5\eta - r_t(j)) \leq \frac{S_j \ln T}{\eta} + \eta \sum_{t=1}^{T} a_t(j)(5\eta - r_t(j))^2$$

$$\leq \frac{S_j \ln T}{\eta} + 4\eta \sum_{t=1}^{T} a_t(j),$$

where the last step uses the fact $5\eta - r_t(j) \in [-1, 2]$. Rearranging completes the proof. $\square$

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

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

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

Proof. For each $j \in U$, combining the previous two results shows

$$\sum_{t:j_t=j} r_t(j) \leq \frac{\ln N}{\eta} + \frac{S_j \ln T}{\eta} + 5\eta T_j.$$

Summing over all $j \in U$, we obtain

$$\mathcal{R}_T(j_1, \ldots, j_T) = \sum_{j \in U} \sum_{t:j_t=j} r_t(j) \leq \frac{n \ln N}{\eta} + \frac{\sum_{j \in U} S_j \ln T}{\eta} + 5\eta \sum_{j \in U} T_j$$

$$\leq \frac{n \ln N}{\eta} + \frac{3S \ln T}{\eta} + 5\eta T,$$

where the last step uses the facts $\sum_{j \in U} T_j = T$ and $\sum_{j \in U} S_j \leq 2S + n \leq 3S$ (since every switch from j to j' in the sequence $j_1, \ldots, j_T$ contributes one switch to S_j and another switch to $S_{j'}$). Picking the order-optimal $\eta = \sqrt{(S \ln T + n \ln N)/T}$ then finishes the proof. $\square$