SUFE: Algorithmic Game Theory Summer School	Problem Set 2a
Due: July 28, 2017, 23:59pm	*Submit solutions to as many problems as you can.*

Please send your solutions to **nikolai@mail.shufe.edu.cn**. You may submit typed in latex pdfs, or scans/photos of your handwritten solutions.

1. **Convex Zero-Sum Games**: Consider the generalization of zero sum games where player 1 chooses a strategy x in some convex set $S \subseteq R^d$ and player 2 chooses a strategy y in some convex set $T \subseteq R^d$. The loss of player 1 is $c(x, y)$ and the loss of player 2 is $-c(x, y)$ where $c(\cdot, \cdot)$ is a function that is convex in its first argument and concave in its second argument.

   a. Show that if the game is played repeatedly and each player employs an online learning algorithm with $\epsilon(T)$-regret, then their average strategies constitute an $O(\epsilon(T))$-approximate Nash equilibrium.

   b. Show the general version of von-Neumman's minimax theorem: i.e.

$$\min_{x\in S}\max_{y\in T} c(x, y) = \max_{y\in T}\min_{x\in S} c(x, y) \tag{1}$$

2. **Optimal Regret Bounds**: The goal of this problem is to show that there exists no online learning algorithm choosing between two actions, whose regret in T rounds is $o(\sqrt{T})$.

   Hint: You can use without proof the fact that an unbiased random walk that goes 1 *step left with probability* 1/2 *and* 1 *step right with probability* 1/2 *will be at distance at least* $\sqrt{T}/2$ *from the origin with probability at least* 1/2*, for large enough* T.

3. **Follow the Perturbed Leader:** Consider the *experts* online learning setting, where at every iteration t the learner picks an action i_t from among K actions, letting $[K]$ denote the set of actions. The adversary picks a loss $\ell_t^i \in [0, 1]$, for each action $i \in [K]$, and the player receives a loss of $\ell_t^{i_t}$. In class, we saw that the Follow-the-Regularized-Leader algorithm, which picks an action at time-step t drawn from a probability distribution $p_t \in \Delta_K$ chosen according to the equation:

$$p_t = \arg\min_{p\in\Delta_K} \sum_{\tau=1}^{t-1} \langle p, \ell_t^i\rangle + \frac{1}{\eta} R(p), \tag{2}$$

   where $R(\cdot)$ is a 1-strongly convex function and η is an appropriately chosen constant of order $O(1/\sqrt{T})$, has expected regret $O(\sqrt{T})$.

   We now consider a slightly different algorithm: (i) Independently at each time-step t, the learner draws a random vector $\epsilon_t = (\epsilon_t^1, \ldots, \epsilon_t^K)$, where each coordinate ϵ_t^j is drawn independently from a uniform distribution supported on $[0, 1/\eta]$. Let D denote the distribution of the vector ϵ_t. (ii) After drawing ϵ_t, the learner adds this vector to the vector recording every action's cumulative loss until step $t-1$ (inclusive), and picks the coordinate with the smallest value, i.e.:

$$i_t = \arg\min_{i\in[K]} \sum_{\tau=1}^{t-1} \ell_\tau^i + \epsilon_t^i. \tag{3}$$

   We call this algorithm Follow-the-Perturbed-Leader.

   a. Consider first the Be-the-Perturbed-Leader version of the algorithm, where the algorithm picks:

$$i_t^* = \arg\min_{i\in[K]} \sum_{\tau=1}^{t} \ell_\tau^i + \epsilon_t^i. \tag{4}$$

Show that, for all sequences of loss vectors $\ell_1, \ldots, \ell_T$, the expected regret of Be-the-Perturbed-Leader is upper bounded by:

$$E_{\epsilon_1,\ldots,\epsilon_T}\left[\sum_{t=1}^{T} \ell_t^{i_t^*} - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t^i\right] \leq E_{\epsilon\sim D}\left[\max_{i\in[K]} \epsilon^i\right]. \tag{5}$$

b. Show that, for all sequences of loss vectors $\ell_1, \ldots, \ell_T$, the expected regret of Follow-the-Perturbed-Leader is upper bounded by:

$$E_{\epsilon_1,\ldots,\epsilon_T}\left[\sum_{t=1}^{T} \ell_t^{i_t} - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t^i\right] \leq E_{\epsilon\sim D}\left[\max_{i\in[K]} \epsilon^i\right] + \sum_{t=1}^{T} E_{\epsilon_t\sim D}\left[1\{i_t \neq i_t^*\}\right]. \tag{6}$$

c. Show that, for all sequences of loss vectors $\ell_1, \ldots, \ell_T$, for each time-step t:

$$E_{\epsilon_t}\left[1\{i_t \neq i_t^*\}\right] \leq K \cdot \eta. \tag{7}$$

and conclude that with an appropriately chosen η, the algorithm has regret $O(\sqrt{KT})$.