Problem 1. (Coarse) correlated equilibria and no-regret learning (3.5 points)

Consider the 2-player identical interest game below with $0 < \epsilon < 1/2$ and both players being maximizers.

$$\begin{array}{c} \\ A \\ A^- \\ B \\ B^- \end{array} \begin{array}{c} \begin{array}{cccc} A & A^- & B & B^- \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1-\epsilon & 1-\epsilon & -\epsilon & -\epsilon \\ 0 & 0 & 1 & 1 \\ -\epsilon & -\epsilon & 1-\epsilon & 1-\epsilon \end{bmatrix} \end{array}$$

We will use (a_1, a_2) to denote a given strategy a_1 of player 1, and a_2 of player 2.

a) Show that the distribution that puts equal probability on (A^-, A^-), and (B^-, B^-) is a coarse-correlated equilibrium. (1 point)

Solution:
To prove that it is a coarse correlated equilibrium we need to show that no agent can increase her reward by unilaterally changing to any alternative fixed action. First let's compute the reward for the two agents with the given distribution:

$$\mathbb{E}_{a\sim D}[U(a)] = \frac{1}{2}U(A^-, A^-) + \frac{1}{2}U(B^-, B^-) = \frac{1}{2}(1-\epsilon) + \frac{1}{2}(1-\epsilon) = 1-\epsilon.$$

The possible actions for player 1 are:

$$\begin{aligned}
\mathbb{E}_{a\sim D}[U(A, a_{-i})] &= \frac{1}{2}U(A, A^-) + \frac{1}{2}U(A, B^-) = \frac{1}{2} \\
\mathbb{E}_{a\sim D}[U(A^-, a_{-i})] &= \frac{1}{2}U(A^-, A^-) + \frac{1}{2}U(A^-, B^-) = \frac{1}{2}(1-\epsilon) + \frac{1}{2}(-\epsilon) = \frac{1}{2} - \epsilon \\
\mathbb{E}_{a\sim D}[U(B, a_{-i})] &= \frac{1}{2}U(B, A^-) + \frac{1}{2}U(B, B^-) = \frac{1}{2} \\
\mathbb{E}_{a\sim D}[U(B^-, a_{-i})] &= \frac{1}{2}U(B^-, A^-) + \frac{1}{2}U(B^-, B^-) = \frac{1}{2}(-\epsilon) + \frac{1}{2}(1-\epsilon) = \frac{1}{2} - \epsilon
\end{aligned}$$

which are all smaller than $1-\epsilon$. For player 2 we have:

$$\begin{aligned}
\mathbb{E}_{a\sim D}[U(A, a_{-i})] = \mathbb{E}_{a\sim D}[U(A^-, a_{-i})] &= \frac{1}{2}U(A^-, A) + \frac{1}{2}U(B^-, A) = \frac{1}{2}(1-\epsilon) + \frac{1}{2}(-\epsilon) = \frac{1}{2} - \epsilon \\
\mathbb{E}_{a\sim D}[U(B, a_{-i})] = \mathbb{E}_{a\sim D}[U(B^-, a_{-i})] &= \frac{1}{2}U(A^-, B) + \frac{1}{2}U(B^-, B) = \frac{1}{2}(-\epsilon) + \frac{1}{2}(1-\epsilon) = \frac{1}{2} - \epsilon
\end{aligned}$$

which are smaller than $1-\epsilon$ as well. Thus, no agent has an incentive in switching to a fixed action.

b) Show that the above distribution is not a correlated equilibrium. (.5 point)

Solution:
To prove that it is not a correlated equilibrium we need to show that, for at least one of the two players, there exists an action $\tilde{a}_i$ such that:

$$\mathbb{E}_{a\sim D}[U(\tilde{a}_i, a_{-i})|a_i] > \mathbb{E}_{a\sim D}[U(a)|a_i].$$

Notice that if player 1 plays A instead of A^- (or B instead of B^-), the reward would be:

$$\mathbb{E}_{a\sim D}[U(A, A^-)|A^-] = 1 > 1-\epsilon = \mathbb{E}_{a\sim D}[U(A^-, A^-)|A^-].$$

Thus the distribution proposed is not a correlated equilibrium.

c) Suppose players are playing this game repeatedly. The empirical sequence of played actions for player 1 over 10 rounds of the algorithm is given as $\{A, A^-, B, B^-, A, A, B^-, B, B^-, A\}$. What is the empirical probability of playing action A for player 1? (.5 point)

Solution:
Player 1 played the action A four times over 10 rounds, thus the empirical probability is:

$$\frac{4}{10} = 0.4 = 40\%.$$

d) Let σ_t be the the empirical probability of the played actions for both players implementing the multiplicative weight update algorithm. Can σ_t converge to (A,A) as $t \to \infty$? Justify. (.5 point)

Solution:
The algorithm is a no-regret algorithm and thus converges to a coarse correlated equilibrium. The algorithm can converge to (A,A), because it is a Nash equilibrium and thus also a coarse correlated equilibrium. However, since it is not the only one, it could also converge to other policies.

e) For a given strategy $b_t = B$ of player 2, write the weight update equation for player 1's actions at $t + 1$, corresponding to the multiplicative weight update algorithm. Hint: to cast the problem as minimization with losses in $[0, 1]$, define the cost $C(a_1, a_2) = 1 - J(a_1, a_2)$, where $J(a_1, a_2)$ is given by matrix above. (1 point)

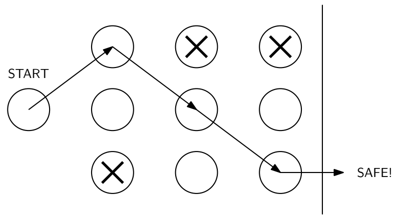
Solution:
Given the strategy of player 2, the cost for player 1 is:

$$C_t(a_1, B) = \frac{1}{1+\epsilon}\left(\begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \begin{bmatrix}U(A,B)\\U(A^-,B)\\U(B,B)\\U(B^-,B)\end{bmatrix}\right) = \frac{1}{1+\epsilon}\left(\begin{bmatrix}1-0\\1-(-\epsilon)\\1-1\\1-(1-\epsilon)\end{bmatrix}\right) = \begin{bmatrix}\frac{1}{1+\epsilon}\\1\\0\\\frac{\epsilon}{1+\epsilon}\end{bmatrix}.$$

We had to multiply the cost for $\frac{1}{1+\epsilon}$ to maintain the cost in the range $[0, 1]$. The weights update is:

$$w_{t+1}(a_1, B) = w_t(a_1, B)(1-\beta)^{c_t(a)} = \begin{bmatrix}w_t(A,B)(1-\beta)^{\frac{1}{1+\epsilon}}\\w_t(A^-,B)(1-\beta)\\w_t(B,B)\\w_t(B^-,B)(1-\beta)^{\frac{\epsilon}{1+\epsilon}}\end{bmatrix}$$

Problem 2. Dynamic programming for zero-sum games (6.5 points)



Recall our escape game from Quiz 1: Player 1 (Alice) is trying to escape, going from the **start node** to the **safe zone** without being intercepted. At every stage of the game, Alice moves one step closer to the safe zone. She can decide to continue on the same row, or instead move one row up (if not in the top row) or one row down (if not in the bottom row). Player 2 (Eve) is trying to stop Alice. At each stage, Eve is aware of Alice's current position, and she is allowed to block one of the three rows in the next stage, taking her decision simultaneously with Alice. If she selects the row corresponding to Alice's next move, the game ends: Eve wins the game getting a cost of -1. Otherwise, the game continues to the next stage. If Alice reaches the safe zone, Alice wins the game and gets a cost of -1. The game is zero-sum.

Let K denote the number of stages of the game, that is, the number of times Alice needs to not get intercepted to reach the safe zone. Notice that $n_K = 1 + 3^2 + 3^4 + \cdots + 3^K$ is the upper bound for the number of information sets for either player.

a) What is the number of decision variables of the linear program for finding a mixed strategy Nash equilibrium? (.5 point)

Solution:
As we have a zero-sum game, we can formulate a linear program with n_K decision variables to determine the mixed Nash equilibrium.

b) How many linear programs you'd have to solve to find a behavioral strategy subgame perfect equilibrium? (.5 point)

Solution:
A behavioral strategy assigns a probability distribution to each information set over the set of possible actions. We have to solve one linear program for every information set, thus no more than n_K linear programs, each of size 3.

As you see from above for K getting large, finding a Nash equilibrium using a game tree is ineffective - we will have to either solve one very large (exponential in K number of variables) linear program to find the mixed strategy Nash equilibrium or many (exponentially in K) small (3 variables) linear programs to find the behavioral strategy subgame perfect Nash equilibrium. We are going to solve the problem with linearly (in K) many linear programs of size 3, a huge improvement! Intuitively, it will be sufficient to consider a dynamic game with states corresponding to Alice's position at each stage (see the figure).

Let us use define the state by $x \in X := \{1, 2, 3, 4\}$, where 1, 2, 3 denote Alice being in the up, middle or bottom positions, respectively. State 4 denotes Alice being intercepted and the game ending. Alice and Eve's action sets are $A = B = \{U, M, D\}$ for choosing up, middle, or bottom.

c) Define the dynamics of the game by constructig the function $f : (x, a, b) \mapsto X$ with $x \in X$, $a \in A$, $b \in B$. Hints: the dynamics from states 4 and 1 are derived below. (1.5 points)

$f(4, a, b) = 4$ for any $a \in A, b \in B$

$f(1, U, U) = f(1, M, M) = 4, \quad f(1, U, M) = f(1, U, D) = 1, \quad f(1, M, U) = 2, \quad f(1, M, D) = 2$

$f(2, U, U) = f(2, M, M) = f(2, D, D) = \ldots$

$f(2, U, M) = \ldots, \quad f(2, U, D) = \ldots, \quad f(2, M, U) = \ldots, \quad f(2, M, D) = \ldots, \quad f(2, D, U) = \ldots, \quad f(2, D, M) = \ldots$

$f(3, D, D) = f(3, M, M) = \ldots$

$f(3, M, U) = \ldots, \quad f(3, M, D) = \ldots, \quad f(3, D, U) = \ldots, \quad f(3, D, M) = \ldots \quad .$

Solution:

$f(2, U, U) = f(2, M, M) = f(2, D, D) = 4$

$f(2, U, M) = 1, \quad f(2, U, D) = 1, \quad f(2, M, U) = 2, \quad f(2, M, D) = 2, \quad f(2, D, U) = 3, \quad f(2, D, M) = 3$

$f(3, D, D) = f(3, M, M) = 4$

$f(3, M, U) = 2, \quad f(3, M, D) = 2, \quad f(3, D, U) = 3, \quad f(3, D, M) = 3 \quad .$

d) Note that at the final stage K, from any position $x \in \{1, 2, 3\}$, Alice moves to the safe zone. Hence, define $V_K(x) = -1$ for $x \in \{1, 2, 3\}$ and $V_K(4) = 1$. Now, write the condition for $V_{K-1}(x)$ to be the value of the game at stage $K - 1$ and $\gamma^*_{K-1} : X \to \mathcal{Y}, \sigma^*_{K-1} : X \to \mathcal{Z}$ to be a corresponding mixed Nash strategy, with $\mathcal{Y}, \mathcal{Z}$ denoting simplexes of appropriate dimension. Hint: there is no stage cost in this game. (1 point)

Solution: The conditions are:

$$V_{K-1}\left(f(x, \gamma^*_{K-1}, \sigma_{K-1})\right) \leq V_{K-1}\left(f(x, \gamma^*_{K-1}, \sigma^*_{K-1})\right) \leq V_{K-1}\left(f(x, \gamma_{K-1}, \sigma^*_{K-1})\right), \quad \forall \gamma_{K-1}, \sigma_{K-1},$$

and

$$\begin{aligned} V_{K-1}\left(f(x, \gamma^*_{K-1}, \sigma^*_{K-1})\right) &= \min_{\gamma_{K-1}} \max_{\sigma_{K-1}} V_{K-1}\left(f(x, \gamma_{K-1}, \sigma_{K-1})\right) \\ &= \max_{\sigma_{K-1}} \min_{\gamma_{K-1}} V_{K-1}\left(f(x, \gamma_{K-1}, \sigma_{K-1})\right). \end{aligned}$$

e) Compute $V_{K-1}(2)$. Hint: You need to formulate a matrix game with 3 rows and 3 columns, and values of $V_K(f(2, a, b))$ are the entries of the matrix, with a, b chosen appropriately. (1 point)

Solution:
The reward matrix is:

$$V_K(f(2, a, b)) = \begin{bmatrix} V_K(f(2, U, U) & V_K(f(2, U, M) & V_K(f(2, U, D) \\ V_K(f(2, M, U) & V_K(f(2, M, M) & V_K(f(2, M, D) \\ V_K(f(2, D, U) & V_K(f(2, D, M) & V_K(f(2, D, D) \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

At the Nash equilibrium, the strategy $[y_1, y_2, y_3]$ for Alice has to be such that:

$$\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} q^* \\ q^* \\ q^* \end{bmatrix}.$$

By solving that linear program we find that $y_1 = y_2 = y_3$. Since their sum needs to be equal to 1, the Nash equilibrium strategy for Alice is $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$. Thus

$$V_{K-1}(2) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = -\frac{1}{3}.$$

f) Assuming you have computed $V_{K-1}(x)$ for $x \in X$ as above, now formulate the matrix game you'd have to solve to find $V_{K-2}(1)$. (.5 point)

Solution:
Alice can play U and M, while Eve can play all three actions (even if D is obviously not convenient, since by playing that action she has zero probability of catching Alice). The reward matrix is:

$$\begin{bmatrix} V_{K-1}\left(f(1,U,U)\right) & V_{K-1}\left(f(1,U,M)\right) & V_{K-1}\left(f(1,U,D)\right) \\ V_{K-1}\left(f(1,D,U)\right) & V_{K-1}\left(f(1,M,M)\right) & V_{K-1}\left(f(1,M,D)\right) \end{bmatrix} = \begin{bmatrix} 1 & V_{K-1}(1) & V_{K-1}(1) \\ V_{K-1}(2) & 1 & V_{K-1}(2) \end{bmatrix}$$

g) How many linear programs of size maximum 3 you'd have to solve to find the value of the game, $V_0(x)$, for $K = 10$? (.5 point)

Solution:
For each state $x \in \{1, 2, 3\}$ we need to solve one linear program. In total we will have $3 \times K = 30$ linear programs.

h) Suppose Alice starts playing the game with $K = 10$ stages, and finds herself at state 3 in stage $t = 5$. With what probabilities should she choose actions M and D? You don't need to give numerical values, only the symbol. (.5 point)

Solution:
To determine the optimal strategy we need to solve a linear program. Being $[y^*, 1-y^*]$ the strategy of Alice and $[z^*, 1-z^*]$ the strategy of Eve (we discard the possibility of choosing U, because it can not lead to catch Alice, thus should not be played), we have:

$$\begin{aligned} V_5(3) &= \begin{bmatrix} y^* & 1-y^* \end{bmatrix} \begin{bmatrix} V_6\left(f(3,M,M)\right) & V_6\left(f(3,M,D)\right) \\ V_6\left(f(3,D,M)\right) & V_6\left(f(3,D,D)\right) \end{bmatrix} \begin{bmatrix} z^* \\ 1-z^* \end{bmatrix} \\ &= \begin{bmatrix} y^* & 1-y^* \end{bmatrix} \begin{bmatrix} 1 & V_6(2) \\ V_6(3) & 1 \end{bmatrix} \begin{bmatrix} z^* \\ 1-z^* \end{bmatrix}. \end{aligned}$$

For the strategy $[y^*, 1-y^*]$ to be otpimal, it needs to hold that:

$$\begin{bmatrix} y^* & 1-y^* \end{bmatrix} \begin{bmatrix} 1 & V_6(2) \\ V_6(3) & 1 \end{bmatrix} = \begin{bmatrix} q^* \\ q^* \end{bmatrix}.$$

Thus we obtain:

$$y^* + (1 - y^*)V_6(3) = y^* V_6(2) + 1 - y^*,$$

which leads to:

$$y^* = \frac{1 - V_6(3)}{2 - V_6(3) - V_6(2)}.$$

So Alice should play M with probability y^* and D with probability $1 - y^*$.