Problem 1. Leader-follower game

a) Consider the two autonomous driving game from Lecture 1.

(i) What is the security strategy of player 1? What would be the game outcome if both players play their corresponding security strategies?

Solution: The matrix of the payoffs is the following:

	Remain	Swerve
Remain	$(30, 0)$	$(30, 10)$
Swerve	$(100, 100)$	$(0, 10)$

The security strategy of a player i is the action that minimizes the worst case scenario. For every action of that player, we see what is the worst possible cost and then we choose the action that minimizes it. For agent 1, the worst possible cost if it plays the action *remain* is 30, while if it plays *swerve*, the worst cost is 100 (in the case that the other agent plays *remain*). Thus, the security strategy for player 1 is to play *remain*. For player 2, the worst cost is 100 if it plays *remain*, and 10 if it plays *swerve*. Thus, its security strategy is *swerve*. If both players play their security strategy, the game outcome would be $(30, 10)$.

(ii) Compute the Stackelberg equilibria for two cases: player 1 being the leader, and player 2 being the leader. Verify that if either of the players act as a leader, its payoff won't be worse than that of her/his Nash equilibrium.

Solution: In the case that player 1 is the leader we obtain the following possibilities:

a) Player 1 announces that it decides to *remain*:
   a) Player 2 plays *remain*. The final cost is $(30, 0)$.
   b) Player 2 plays *swerve*. The final cost is $(30, 10)$.

   For Player 2 the optimal choice is to *remain*, so its cost will be 0. The cost for Player 1 will be 30.
b) Player 1 announces that it decides to *swerve*:
   a) Player 2 plays *remain*. The final cost is $(100, 100)$.
   b) Player 2 plays *swerve*. The final cost is $(0, 10)$.

   For Player 2 the optimal choice is to *swerve*, so its cost will be 10. The cost for Player 1 will be 0.

The optimal strategy for player 1 is to announce that it is going to play *swerve*, and the optimal strategy for player 2 is to play *swerve* as well. The cost will be $(0, 10)$. In the case that player 2 is the leader we obtain the following possibilities:

a) Player 2 announces that it decides to *remain*:
   a) Player 1 plays *remain*. The final cost is $(30, 0)$.
   b) Player 1 plays *swerve*. The final cost is $(100, 100)$.

   For player 1 the optimal choice is to *remain*, so its cost will be 30. The cost for player 2 will be 0.
b) Player 2 announces that it decides to *swerve*:
   a) Player 1 plays *remain*. The final cost is $(30, 10)$.
   b) Player 1 plays *swerve*. The final cost is $(0, 10)$.

   For player 1 the optimal choice is to *swerve*, so its cost will be 0. The cost for player 2 will be 10.

The optimal strategy for Player 2 is to announce that it is going to play *remain*. The optimal strategy for player 1 is then to play *remain*. The cost will be $(30, 0)$. There are two Nash equilibria in this game: (*remain*, *remain*) and (*swerve*, *swerve*). Note that the game has a pure strategy Nash equilibrium and if player i is the leader, rational reaction set of the other player is a singleton for i = 1,2. Hence, assumptions of the theorem on slide 18 of Lecture 05 are satisfied. As proven in this theorem, the leader should not be worse off playing its Stackelberg equilibrium strategy than its Nash equilibrium. This is true, since the strategy we have found correspond to the two Nash equilibria.

b) Now consider a zero-sum finite action game where there is no pure strategy Nash equilibrium. Prove that in this setting, the leader will always be worse off. (Note: in Lecture 05, we saw that in this case, the leader may no longer be better off playing its Stackelberg equilibrium strategy in comparison to its Nash equilibrium strategy as at least one of the assumptions of the theorem is not satisfied).

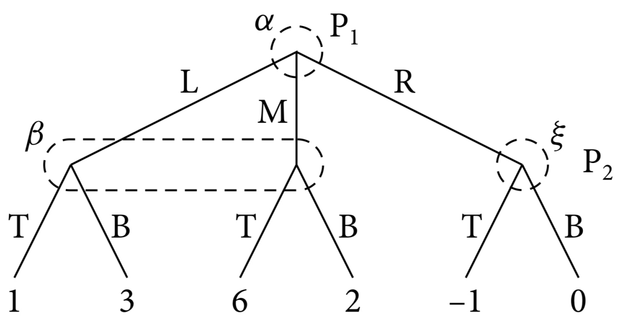
Solution: Let us consider player 1 (the minimizer) as a leader without loss of generality. Then, since there is no pure strategy Nash equilibrium, we know that

$$\overbrace{\underline{V} := \max_{j\in\{1,\ldots,m\}} \min_{i\in\{1,\ldots,n\}} a_{ij}}^{P_2} < \overbrace{\overline{V} := \min_{i\in\{1,\ldots,n\}} \max_{j\in\{1,\ldots,m\}} a_{ij}}^{P_1}$$

Now, observe that for player 1 acting as a leader, $\overline{V}$ is the Stackelberg value of the game. The other player, which is the maximizer, chooses the action after player 1. If player 1 plays action $\bar{i}$, player 2 will play action $\arg\max_{j\in\{1,\ldots,m\}} a_{\bar{i}j}$. Knowing this, player 1 chooses the action $\arg\min_{i\in\{1,\ldots,n\}} \arg\max_{j\in\{1,\ldots,m\}} a_{ij}$. This leads to the value $\overline{V}$. If player 2 acts a leader, then $\underline{V}$ is the Stackelberg value of the game.

Problem 2. Subgame perfect behavioral equilibrium

Consider the feedback game below (see also Figure 7.4 of Hespanha book). Player 1 is a minimizer and player 2 is a maximizer.



a) Is this a perfect information game?

Solution: This is not a perfect information game because the information set β of player 2 contains two nodes. In other words, player 2 cannot distinguish between her first and her second node.

b) Formulate the game in matrix form. Verify that the game has several pure strategy Nash equilibria, find these equilibria. And verify that no pure strategy subgame perfect equilibrium exist.

Solution:

	$\sigma(\beta)=T,\sigma(\xi)=T$	$\sigma(\beta)=T,\sigma(\xi)=B$	$\sigma(\beta)=B,\sigma(\xi)=T$	$\sigma(\beta)=B,\sigma(\xi)=B$
$\gamma(\alpha)=L$	1	1	3	3
$\gamma(\alpha)=M$	6	6	2	2
$\gamma(\alpha)=R$	-1	0	-1	0

Note that the actions L and M are strictly dominated by action R. Thus, the game has the following two pure

Nash equilibria:

$$\text{First one: } P_1 \text{ policy} = \begin{cases} R, & \text{if IS} = \alpha \end{cases},$$

$$P_2 \text{ policy} = \begin{cases} T, & \text{if IS} = \beta \\ B, & \text{if IS} = \xi \end{cases}$$

$$\text{Second one: } P_1 \text{ policy} = \begin{cases} R, & \text{if IS} = \alpha, \end{cases}$$

$$P_2 \text{ policy} = \begin{cases} B, & \text{if IS} = \beta \\ B, & \text{if IS} = \xi \end{cases}$$

The game has the following two subgames:

$$A_\beta = \begin{matrix} & T & B \\ L & 1 & 3 \\ M & 6 & 2 \end{matrix}$$

$$A_\xi = \begin{matrix} & T & B \\ R & -1 & 0 \end{matrix}$$

The game has no pure strategy subgame perfect equilibrium because subgame A_β has no pure Nash equilibrium.

c) Determine the behavioral strategy subgame perfect equilibrium of the game.

Solution: Consider the two subgames:

$$A_\beta = \begin{matrix} & T & B \\ L & 1 & 3 \\ M & 6 & 2 \end{matrix}$$

$$A_\xi = \begin{matrix} & T & B \\ R & -1 & 0 \end{matrix}$$

Subgame A_β has the mixed NE $y^*_\beta = (2/3, 1/3)$ and $z^*_\beta = (1/6, 5/6)$ and the value of the game is $V(A_\beta) = 8/3$. Subgame A_ξ has the mixed NE $y^*_\xi = 1$ and $z^*_\xi = (0, 1)$ and the value of the game is $V(A_\xi) = 0$. Note that for player 1 $V(A_\xi) = 0 < 8/3 = V(A_\beta)$. Thus, the behavioral strategy subgame perfect equilibrium of the game is:

$$P_2 \text{ policy} = \begin{cases} (\frac{1}{6}, \frac{5}{6}), & \text{if IS} = \beta \\ (0, 1), & \text{if IS} = \xi \end{cases}$$

$$P_1 \text{ policy} = \begin{cases} (0, 0, 1), & \text{if IS} = \alpha. \end{cases}$$

d) Verify that the value of the game corresponding to the behavioral strategy subgame perfect equilibrium is the same as the value of any of the pure strategy Nash equilibria. Could you have made this conclusion without computing the equilibria of the game?

Solution: The value of the game corresponding to the behavioral strategy subgame perfect equilibrium is 0 (see c)) and the value of the two NEs is also zero (look at the game in matrix form in b)).

From slide 26 of Lecture 06, we know for any feedback game, a Nash equilibrium in behavioral strategies has the same value as any mixed strategy Nash equilibrium. This is a feedback game, and hence, the statement holds. Furthermore, since it is a zero-sum game, all equilibria (mixed/pure) have the same value. Hence, the statement above holds.

Note that the proof of the statement on slide 26 is based on the so-called Kuhn's theorem, which we did not discuss in the course.

Problem 3. Correlated equilibria and no-regret learning

Consider the identical interest game where players are utility maximizers:

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 2 & 1 & -4 \\ 1 & 0 & -1 \\ -4 & -1 & -2 \end{bmatrix} \end{array}$$

This game is taken from Figure 3 (i) of the following article [1]. The row player corresponds to player 1 and the column player corresponds to player 2.

a) Find the unique pure strategy Nash equilibrium of the game by removing strictly dominated actions.

Solution: For player 1 action C is strictly dominated by action B, so we can remove row 3. Similarly, for player 2 action C is strictly dominated by action B, so we can remove column C. We get that (A, A) is the NE with players' utilities (2,2).

b) Verify that the probability distribution D that assigns probability 1/3 to strategies corresponding to the diagonal entries is a coarse correlated equilibrium (CCE) of the game.

Solution: In the following, we consider player 1 but the argument holds also for player 2 since the game is identical interest. We calculate the expected payoff of player 1 if this player samples her action from the CCE:

$$\mathbb{E}_{a\sim D}[U^1(a)] = \frac{1}{3} * 2 + \frac{1}{3} * 0 + \frac{1}{3} * (-2) = 0.$$

Suppose player 2 continues to play according to the CCE. If player 1 decides to deviate to A, her expected payoff is:

$$\mathbb{E}_{a\sim D}[U^1(A, a_2)] = \frac{1}{3} * 2 + \frac{1}{3} * 1 + \frac{1}{3} * (-4) = -\frac{1}{3}.$$

If player 1 decides to deviate to B, her expected payoff is:

$$\mathbb{E}_{a\sim D}[U^1(B, a_2)] = \frac{1}{3} * 1 + \frac{1}{3} * 0 + \frac{1}{3} * (-1) = 0.$$

If player 1 decides to deviate to B, her expected payoff is:

$$\mathbb{E}_{a\sim D}[U^1(C, a_2)] = \frac{1}{3} * (-4) + \frac{1}{3} * (-1) + \frac{1}{3} * (-2) = -\frac{7}{3}.$$

Hence, player 1 cannot improve her expected payoff by unilateral deviation.

c) According to Lecture 01, rational players should not choose strictly dominated actions. Argue that a coarse correlated equilibrium may not satisfy the rationality model above.

Solution: In a) we saw that (C, C) is a strictly dominated action profile but the CCE puts weight $1/3$ on it. Hence, this is an example showing that CCEs may also have undesirable properties if we were to accept the rationality model of players we assumed when defining a Nash equilibrium.

d) Now, we consider player 1 implementing the multiplicative weight update algorithm, while player 2 plays her constant Nash equilibrium strategy. Let the sequence of played actions by player 1 be $\{a_t^1\}_{t=1}^T$. Based on the no-regret property of this algorithm, what should the empirical frequency $\sigma_\tau^1(a) = \frac{1}{\tau}\{|t \in [\tau] : a = a_t^1\}$, $\tau \in [T]$ of player 1 converge to?

Solution: The action of player 2 is fixed to A. Then, the utility of player 1 is given by $U^1(A) = U^1(A, A) = 2$, $U^1(B) = U^1(B, A) = 1$, and $U^1(C) = U^1(C, A) = -4$. Recall that for games with utilities, regret of player i is defined as follows:

$$R^i(\tau) = \frac{1}{\tau}\left(\max_a \sum_{t=1}^{\tau} U^i(a, a_t^{-i}) - \sum_{t=1}^{\tau} U^i(a_t^i, a_t^{-i})\right), \quad \tau \in [T].$$

Since $a_t^2 = A$ for all $t \in [\tau]$ it follows that $A = \arg\max_a \sum_{t=1}^{\tau} U^1(a, A)$. To ensure the no-regret property of the multiplicative weight update algorithm, i.e. $\lim_{\tau\to\infty} R^i(\tau) = 0$, player 1 must play her unique best response to player 2 playing A, namely play A as well. Thus, $\lim_{\tau\to\infty} \sigma_\tau^1(a) = 1_{\{a=A\}}$.

e) Now, consider both players implementing the multiplicative weight update algorithm. What should the empirical frequency $\sigma_\tau(a) = \frac{1}{\tau}\{|t \in [\tau] : a_t = (a_t^1, a_t^2)\}$, $\tau \in [T]$ converge to? Can you conclude anything about the convergence of the actual sequence of played actions $\{a_t\}_{t=1}^T$?

Solution: The empirical frequency $\sigma_\tau(a) = \frac{1}{\tau}\{|t \in [\tau] : a_t = (a_t^1, a_t^2)\}$, $\tau \in [T]$ converges to a CCE in general. Note that a pure strategy Nash equilibrium, a mixed strategy Nash equilibrium, and a correlated equilibrium are CCEs. We cannot tell (based on what we learned so far) to which one of the CCEs a no-regret learning algorithm such as the multiplicative weights algorithm will converge to.

Problem 4. Dynamic games: Zero-sum LQR

In Lecture 02, we proved that a zero-sum game has a mixed strategy Nash equilibrium, also referred to as saddle point equilibrium, and the game has a value. To this end, we leveraged our Lecture 01 result on the existence of mixed strategy Nash equilibria for general-sum games. Recall that a saddle point equilibrium has the property that

$$\max_{z\in\mathcal{Z}} \min_{y\in\mathcal{Y}} \; y^\top A z = \min_{y\in\mathcal{Y}} \max_{z\in\mathcal{Z}} \; y^\top A z,$$

with $\mathcal{Y} \subset \mathbb{R}^{m_1}$, $\mathcal{Z} \subset \mathbb{R}^{m_2}$ being simplexes and m_i denoting the number of actions of player i, $i \in \{1, 2\}$.

A more general result from convex optimization is that if a function $J : Y \times Z \to \mathbb{R}$ is convex in y and concave in z, with Y, Z being convex sets, we have:

$$\min_{y\in Y} \max_{z\in Z} J(y, z) = \max_{z} \min_{y} J(y, z). \tag{0.1}$$

Based on the above result, verify the stated result in Exercise 4.3 of Hespanha. Next, equipped with the above, we will tackle zero-sum linear quadratic games.

Consider a zero-sum discrete-time linear quadratic game with state dynamics for $k = 1, 2, ...$ given as

$$x_{k+1} = Ax_k + Bu_k + Ev_k,\; x \in \mathbb{R}^n, u \in \mathbb{R}^{m_u}, v \in \mathbb{R}^{m_v}.$$

Above, u_k, v_k are player 1 and 2 actions at time k, respectively. Player 1 aims to minimize the following cost function while player 2 aims to maximize it:

$$\sum_{k=1}^{K} \left(x_k^T Q x_k + u_k^T u_k \right) - \mu^2 v_k^T v_k.$$

Above, $Q \succeq 0$, and thus, $Q = C^T C$ for some $C \in \mathbb{R}^{p\times n}$, $p \leq n$ (from properties of positive semidefinite matrices). Furthermore, μ^2 can be considered as a constant that maps units of v_k to that of u_k (if players' have different actuation possibilities). The control interpretation of the formulation above is that player 1 aims to keep the output $y_k = Cx_k$ small with minimum energy (thus, penalizing $\|u_k\|_2^2$, while player 2 aims to keep the output large with minimum energy (note that player 2 is maximizing the cost).

a) Consider a pair of strategies (γ, σ), where $\gamma = (\gamma_1, ..., \gamma_K)$, $\sigma = (\sigma_1, ..., \sigma_K)$, $\gamma_t : X \to \mathbb{R}^{m_u}$, $\sigma_t : X \to \mathbb{R}^{m_v}$, $t = 1, 2, ..., K$. Use backward induction to write the equations which a pure strategy subgame perfect equilibrium (γ^*, σ^*) must satisfy. *Hint:* Start the backward recursion with $V_{K+1}(x) = 0$.
Solution: The starting point is, as said in the hint, $V_{K+1}(x) = 0$. We can obtain the other cost-to-go by backward induction with the following formula, assuming for now that σ^* is known:

$$V_k(x) = \inf_{u_k\in U_k} [g_k(x, u_k, \sigma_k^*(x)) + V_{k+1}(f_k(x, u_k, \sigma_k^*(x)))]$$

for all $k \in \{1, 2, \dots, K\}$, and then the pure strategy perfect equilibrium $\gamma^\star$ would be

$$\gamma_k^*(x) := \arg\min_{u_k \in U_k} [g_k(x, u_k, \sigma_k^*(x)) + V_{k+1}(f_k(x, u_k, \sigma_k^*(x)))]$$

for all $k \in \{1, 2, \dots, K\}$. In particular, for the zero-sum LQR we obtain:

$$\begin{aligned} V_k(x) :&= \min_{u_k \in \mathbb{R}^{m_u}} \sup_{v_k \in \mathbb{R}^{m_v}} \left[x^\top C^\top C x + u_k^\top u_k - \mu^2 v_k^\top v_k + V_{k+1}(Ax + Bu_k + Ev_k) \right] \\ &= \max_{v_k \in \mathbb{R}^{m_v}} \inf_{u_k \in \mathbb{R}^{m_u}} \left[x^\top C^\top C x + u_k^\top u_k - \mu^2 v_k^\top v_k + V_{k+1}(Ax + Bu_k + Ev_k) \right] \end{aligned}$$

$\forall x_k \in \mathbb{R}^n$, $k \in \{1, 2, \dots, K\}$ and, inspired by the quadratic form of the stage cost, we will try to find a solution of the form

$$V_k(x) = x^\top P_k x.$$

for appropriately selected symmetric $n \times n$ matrices P_k. Since $V_{K+1}(x) = 0$, we start with $P_{K+1} = 0$, while for P_k with $k \in \{1, 2, \dots, K\}$ we will have

$$x^\top P_k x = \min_{u_k \in \mathbb{R}^{m_u}} \sup_{v_k \in \mathbb{R}^{m_v}} Q_x(u_k, v_k) \tag{0.2}$$

$$= \max_{v_k \in \mathbb{R}^{m_v}} \inf_{u_k \in \mathbb{R}^{m_u}} Q_x(u_k, v_k), \tag{0.3}$$

where

$$Q_x(u_k, v_k) := x^\top C_k x + u_k^\top u_k - \mu^2 v_k^\top v_k + (Ax + Bu_k + Ev_k)^\top P_{k+1}(Ax + Bu_k + Ev_k)$$

b) Under which condition a pure strategy subgame perfect linear state feedback equilibrium exists?
Solution: We can find the Nash equilibrium by computing the optimal $u_k^*(v_K)$ and $v_k^*(u_k)$ and solving the following system:

$$\begin{cases} u_k^{NE} = u_k^*(v_K^{NE}), \\ v_k^{NE} = v_k^*(u_k^{NE}). \end{cases} \tag{0.4}$$

We can compute the optimal $u_k^*(v_K)$ (and $v_k^*(u_k)$) by taking the derivative of $Q_x(u_k, v_k)$ with respect to u_k (and v_k) and putting it equal to zero. In fact, $Q_x(u_k, v_k)$ is a quadratic function with respect to u_k (and v_k), the only stationary point is the optimal policy. By doing this, the system 0.4 becomes

$$\begin{bmatrix} I + B^\top P_{k+1} B & B^\top P_{k+1} E \\ E^T P_{k+1} B & -\mu^2 I + E P_{k+1} E^T \end{bmatrix} \begin{bmatrix} u^* \\ d^* \end{bmatrix} = - \begin{bmatrix} B^\top P_{k+1} A \\ E^\top P_{k+1} A \end{bmatrix} x.$$

This saddle-point equilibrium exists (and it is unique), if the matrix

$$\begin{bmatrix} I + B^\top P_{k+1} B & B^\top P_{k+1} E \\ E^T P_{k+1} B & -\mu^2 I + E P_{k+1} E^T \end{bmatrix}$$

is invertible. For a more detailed explanation look at Chapter 17.4 on Linear Quadratic Dynamic Games in the book Noncooperative Game Theory from Hespanha (link on Moodle).

References

[1] Yannick Viossat and Andriy Zapechelnyuk. No-regret dynamics and fictitious play. *Journal of Economic Theory*, 148(2):825–842, 2013.