COSC-6590/GSCS-6390

Games: Theory and Applications

Lecture 08 - Stochastic Policies for Games in Extensive Form

Luis Rodolfo Garcia Carrillo

School of Engineering and Computing Sciences
Texas A&M University - Corpus Christi, USA

Table of contents

Mixed Policies and Saddle-Point Equilibria

Mixed Policies and Saddle-Point Equilibria

Take a game in extensive form for which

$$\Gamma_1 = \{\gamma_1, \gamma_2, \ldots, \gamma_m\} \qquad \Gamma_2 = \{\sigma_1, \sigma_2, \ldots, \sigma_m\}$$

are the finite sets of all pure policies for P_1 and P_2

Game can be represented by $m \times n$ matrix A_{ext}

- entry a_{ij} is outcome of the game $J(\gamma_i, \sigma_i)$ when P_1 selects the pure policy γ_i (the ith row), P_2 selects the pure policy σ_j (the jth column).

Mixed policy for games in extensive form
Selecting a pure policy randomly according to a previously selected probability distribution before the game starts, and then playing that policy throughout the game.

Mixed Policies and Saddle-Point Equilibria

A **mixed policy** for P_1 is a set of numbers

$$\{y_1, y_2, \dots\ y_m\}, \quad \sum_{i=1}^{m} y_i = 1, \quad y_i \geq 0, \quad \forall i \in \{1, 2, \dots, m\}$$

y_i is the probability that P_1 uses to select the pure policy γ_i.

A **mixed policy** for P_2 is a set of numbers

$$\{z_1, z_2, \dots\ z_n\}, \quad \sum_{j=1}^{n} z_j = 1, \quad z_j \geq 0, \quad \forall j \in \{1, 2, \dots, n\}$$

z_j is the probability that P_2 uses to select the pure policy σ_j.

Mixed Policies and Saddle-Point Equilibria

Assumptions
- random selections by players are statistically independently
- players try to optimize the **expected outcome of the game**

$$J = \sum_{i,j} J(\gamma_i, \sigma_j) \text{Prob}(P_1 \text{ selects } \gamma_i \text{ and } P_2 \text{ selects } \sigma_j) = y' A_{\text{ext}} z,$$

where $y := [y_1 \;\; y_2 \;\; \cdots \;\; y_m]'$ and $z := [z_1 \;\; z_2 \;\; \cdots \;\; z_n]'$.

Use security levels, security policies, and saddle-point equilibria for general games, with the understanding that:

1. the action spaces are the sets $\mathcal{Y}$ and $\mathcal{Z}$ of all mixed policies for players P_1 and P_2
2. for a particular pair of mixed policies $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ the outcome of the game when P_1 uses policy y and P_2 uses policy z is given by $J(y, z) := y' A_{\text{ext}} z$.

Mixed Policies and Saddle-Point Equilibria

Definition 8.1 (Mixed saddle-point equilibrium).

A pair of policies $(y^*, z^*) \in \mathcal{Y} \times \mathcal{Z}$ is **a mixed saddle-point equilibrium** if

$$y^{*\prime} A_{\text{ext}} z^* \leq y' A_{\text{ext}} z^*, \quad \forall y \in \mathcal{Y} \qquad y^{*\prime} A_{\text{ext}} z^* \geq y^{*\prime} A_{\text{ext}} z, \quad \forall z \in \mathcal{Z}$$

and $y^{*\prime} A_{\text{ext}} z^*$ is called the saddle-point value.

Corollary 8.1. For every zero-sum game in extensive form with (finite) matrix representation A_{ext}:

P8.1 A mixed saddle-point equilibrium always exists and

$$\underline{V}_m(A_{\text{ext}}) := \max_{z \in \mathcal{Z}} \min_{y \in \mathcal{Y}} y' A_{\text{ext}} z = \min_{y \in \mathcal{Y}} \max_{z \in \mathcal{Z}} y' A_{\text{ext}} z =: \bar{V}(A_{\text{ext}})$$

Mixed Policies and Saddle-Point Equilibria

P8.2 If y^* and z^* are mixed security policies for P_1 and P_2, then (y^*, z^*) is a mixed saddle-point equilibrium and its value $y^{*\prime} A_{\text{ext}} z^*$ is equal to previous eq.

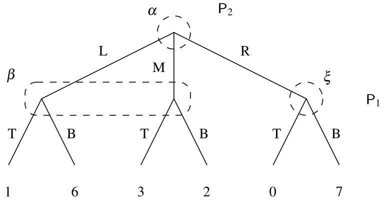
P8.3 If (y^*, z^*) is a mixed saddle-point equilibrium then y^* and z^* are mixed security policies for P_1 and P_2, respectively and previous eq. is equal to the mixed saddle-point value $y^{*\prime} A_{\text{ext}} z^*$.

P8.4 If (y_1^*, z_1^*) and (y_2^*, z_2^*) are mixed saddle-point equilibria then (y_1^*, z_2^*) and (y_2^*, z_1^*) are also mixed saddle-point equilibria and

$$y_1^{*\prime} A_{\text{ext}} z_1^* = y_2^{*\prime} A_{\text{ext}} z_2^* = y_1^{*\prime} A_{\text{ext}} z_2^* = y_2^{*\prime} A_{\text{ext}} z_1^*$$

Mixed Policies and Saddle-Point Equilibria

Example 8.1. Consider the game in extensive form



Enumerate the policies available for P_1 and P_2 as

P_1 :

IS	γ_1	γ_2	γ_3	γ_4
β	T	T	B	B
ξ	T	B	T	B

P_2 :

IS	σ_1	σ_2	σ_3
α	L	M	R

Mixed Policies and Saddle-Point Equilibria

In this case, we obtain

$$A_{\text{ext}} = \underbrace{\left.\begin{bmatrix} 1 & 3 & 0 \\ 1 & 3 & 7 \\ 6 & 2 & 0 \\ 6 & 2 & 7 \end{bmatrix}\right\}}_{\substack{P_2 \text{ policies} \\ (\sigma_1, \sigma_2, \sigma_3)}} \begin{matrix} P_1 \text{ policies} \\ (\gamma_1, \gamma_2, \gamma_3, \gamma_4) \end{matrix}$$

with the following value and security policies

$$V_m(A_{\text{ext}}) = \frac{8}{3} \qquad y^* = \begin{bmatrix} \frac{2}{3} \\ 0 \\ 0 \\ \frac{1}{3} \end{bmatrix} \text{ or } \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{1}{3} \\ 0 \end{bmatrix}, \qquad z^* = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \\ 0 \end{bmatrix}$$

Mixed Policies and Saddle-Point Equilibria

This corresponds to the following choices for the players before the game starts

$$P_1 \textbf{ selects pure policy } \begin{cases} \gamma_1 & \text{with probability } \frac{2}{3} \quad \text{(T)} \\ \gamma_4 & \text{with probability } \frac{1}{3} \quad \text{(B)} \end{cases}$$

or

$$P_1 \textbf{ selects pure policy } \begin{cases} \gamma_1 & \text{with probability } \frac{2}{3} \quad \text{(T)} \\ \gamma_3 & \text{with probability } \frac{1}{3} \quad \text{(B or T)} \end{cases}$$

$$P_2 \textbf{ selects pure policy } \begin{cases} \sigma_1 & \text{with probability } \frac{1}{6} \quad \text{(L)} \\ \gamma_4 & \text{with probability } \frac{5}{6} \quad \text{(M)} \end{cases}$$

Behavioral Policies for Games in Extensive Form

Behavioral Policies for Games in Extensive Form

Behavioral policies involve randomization, done over actions as the game is played, not over pure policies before game starts.

Behavioral policy for P_i: a decision rule π_i (a function) that associates to each information set α of P_i a probability distribution $\pi_i(\alpha)$ over the possible actions for that IS.

- when P_i is in a particular information set α the distribution $\pi_i(\alpha)$ is used to decide on an action for P_i.

Assumptions:

- random selections by both players at the different IS's are all done statistically independently.
- the players try to optimize the resulting expected outcome J of the game.

Behavioral Policies for Games in Extensive Form

As for pure policies, we can perform a per-stage decomposition of behavioral policies for feed-back games.

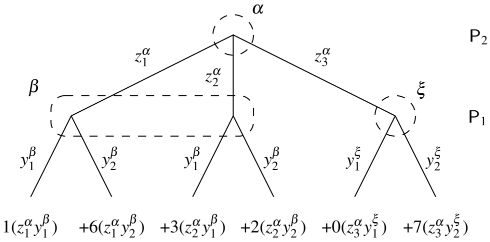
If we are given a behavioral policy γ that maps IS's to probability distributions over actions, we can decompose γ into K sub-policies

$$\gamma := \{\gamma_1, \gamma_2, \ldots, \gamma_K\}$$

where each γ_i only maps the ith stage information sets to probability distributions over actions.

Mixed Policies and Saddle-Point Equilibria

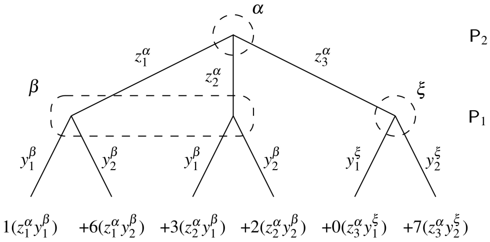
Example 8.2. Consider the game in extensive form



A behavioral policy for P_2 is a function that maps the (only) IS α into a probability distribution as

$$P_2 : \begin{array}{c|ccc} IS & L & M & R \\ \hline \alpha & z_1^\alpha & z_2^\alpha & z_3^\alpha \end{array} \qquad z_1^\alpha, z_2^\alpha, z_3^\alpha \geq 0, \quad z_1^\alpha + z_2^\alpha + z_3^\alpha = 1$$

Mixed Policies and Saddle-Point Equilibria



A behavioral policy for P_1 is a function that maps each of the information sets β , ξ into a probability distribution as follows

P_1 :

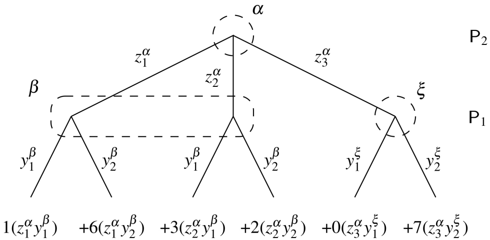
IS	T	B
β	y_1^β	y_2^β
ξ	y_1^ξ	y_2^ξ

$$y_1^\beta, y_2^\beta \geq 0, \quad y_1^\beta + y_2^\beta = 1$$
$$y_1^\xi, y_2^\xi \geq 0, \quad y_1^\xi + y_2^\xi = 1$$

Mixed Policies and Saddle-Point Equilibria

For these behavioral policies, the expected outcome of the game is given by the following expression

$$J = 1z_1^\alpha y_1^\beta + 6z_1^\alpha y_2^\beta + 3z_2^\alpha y_1^\beta + 2z_2^\alpha y_2^\beta + 0z_3^\alpha y_1^\xi + 7z_3^\alpha y_2^\xi$$



Mixed Policies and Saddle-Point Equilibria

Note 9. Procedure to compute the expected outcome for an arbitrary game in extensive form under behavioral policies:

1. Label every link of the tree with the probability with which that action will be chosen by the behavioral policy of the corresponding player.
2. For each leaf multiply all the probabilities of the links that connect the root to that leaf. The resulting number is the probability for that outcome.
3. The expected reward is obtained by summing over all leaves the product of the leaf's outcome by its probability computed in Step 2.

Behavioral Saddle-Point Equilibria

Behavioral Saddle-Point Equilibria

Use the concepts of security levels, security policies, and saddle-point equilibria with the understanding that

1. **action spaces:** the sets $\Gamma_1^\flat$ and $\Gamma_2^\flat$ of all behavioral policies for players P_1 and P_2
2. for a particular pair of behavioral policies $\gamma \in \Gamma_1^\flat$, $\sigma \in \Gamma_2^\flat$, we denote by $J(y, z)$ the **expected outcome of the game** when P_1 uses policy γ and P_2 uses policy σ.

Definition 8.2 (Behavioral saddle-point equilibrium).
A pair of policies $(\gamma^*, \sigma^*) \in \Gamma_1^\flat \times \Gamma_2^\flat$ is a **behavioral saddle-point equilibrium** if

$$J(\gamma^*, \sigma^*) \leq J(\gamma, \sigma^*), \quad \forall \gamma \in \Gamma_1^\flat \qquad J(\gamma^*, \sigma^*) \geq J(\gamma^*, \sigma), \quad \forall \sigma \in \Gamma_2^\flat$$

and $J(\gamma^*, \sigma^*)$ is called the behavioral saddle-point value.

Behavioral Saddle-Point Equilibria

Definition 8.3 (Feedback behavioral saddle-point equlibrium). For a feedback game with K stages, a pair of policies (γ^*, σ^*)

$$\gamma^* := \{\gamma_1^*, \gamma_2^*, \dots, \gamma_K^*\}, \qquad \sigma^* := \{\sigma_1^*, \sigma_2^*, \dots, \sigma_K^*\}$$

is a **feedback behavioral saddle-point equilibrium** if for every stage k and all policies $\gamma_1, \gamma_2, \dots, \gamma_{k-1}$ and $\sigma_1, \sigma_2, \dots, \sigma_{k-1}$ for the stages prior to k we have that

$$J\left(\left\{\gamma_{1\dots k-1}, \boxed{\gamma_k^*}, \gamma_{k+1\dots K}^*\right\}, \left\{\sigma_{1\dots k-1}, \boxed{\sigma_k^*}, \sigma_{k+1\dots K}^*\right\}\right) \leq J\left(\left\{\gamma_{1\dots k-1}, \boxed{\gamma_k}, \gamma_{k+1\dots K}^*\right\}, \left\{\sigma_{1\dots k-1}, \boxed{\sigma_k^*}, \sigma_{k+1\dots K}^*\right\}\right) \quad \forall \gamma_k$$

$$J\left(\left\{\gamma_{1\dots k-1}, \boxed{\gamma_k^*}, \gamma_{k+1\dots K}^*\right\}, \left\{\sigma_{1\dots k-1}, \boxed{\sigma_k^*}, \sigma_{k+1\dots K}^*\right\}\right) \geq J\left(\left\{\gamma_{1\dots k-1}, \boxed{\gamma_k^*}, \gamma_{k+1\dots K}^*\right\}, \left\{\sigma_{1\dots k-1}, \boxed{\sigma_k}, \sigma_{k+1\dots K}^*\right\}\right) \quad \forall \sigma_k$$

where the universal quantifications refer to all possible k stage behavioral sub-policies.

Behavioral Saddle-Point Equilibria

Theorem 8.1. For every zero-sum feedback game G in extensive form with a finite number of stages:

P8.5 A feedback behavioral saddle-point equilibrium always exists and

$$\underline{V}_\flat(G) := \max_{\sigma\in\Gamma_2^\flat} \min_{\gamma\in\Sigma_1^\flat} J(\gamma,\sigma) = \min_{\gamma\in\Sigma_1^\flat} \max_{\sigma\in\Gamma_2^\flat} J(\gamma,\sigma) =: \bar{V}_\flat(G)$$

P8.6 If γ^* and σ^* are behavioral security policies for P_1 and P_2, then (γ^*,σ^*) is a behavioral saddle-point equilibrium and its value $J(\gamma^*,\sigma^*)$ is equal to the previous eq.

Behavioral Saddle-Point Equilibria

P8.7 If (γ^*, σ^*) is a behavioral saddle-point equilibrium, then γ^* and σ^* are behavioral security policies for P_1 and P_2, and the equation is equal to the behavioral saddle-point value $J(\gamma^*, \sigma^*)$.

P8.8 If (γ_1^*, σ_1^*) and (γ_2^*, σ_2^*) are behavioral saddle-point equilibria then (γ_1^*, σ_2^*) and (γ_2^*, σ_1^*) are also behavioral saddle-point equilibria and

$$(\gamma_1^*, \sigma_1^*) = (\gamma_2^*, \sigma_2^*) = (\gamma_1^*, \sigma_2^*) = (\gamma_2^*, \sigma_1^*)$$

Behavioral vs. Mixed Policies

Behavioral vs. Mixed Policies

Total number of distinct **pure policies** for player P_i

$$\underbrace{(\#\text{ actions of 1st IS}) \times (\#\text{ actions of 2nd IS}) \times \cdots (\#\text{ actions of last IS})}_{\text{product over all information sets for } P_i}$$

Then: mixed policies are probability distributions over these many pure actions, which means that we have

$$\underbrace{(\#\text{ actions of 1st IS}) \times (\#\text{ actions of 2nd IS}) \times \cdots \times (\#\text{ actions of last IS})}_{\text{product over all information sets for } P_i} - 1$$

degrees of freedom (DOF) in selecting a mixed policy.

Behavioral vs. Mixed Policies

For **Behavioral Policies**:

For each IS, select a probability distribution over the actions of that IS.

The probability distribution has as many DOF as the # of actions minus one.

Therefore, the total number of DOF available for the selection of behavioral policies is given by

$$\underbrace{(\#\text{ actions of 1st IS} - 1) + (\#\text{ actions of 2nd IS} - 1) + \cdots + (\#\text{ actions of last IS} - 1)}_{\text{sum over all information sets for } P_i}$$

generally a number far smaller than DOF for mixed policies

Behavioral vs. Mixed Policies

By simply counting DOF, we have seen that the set of mixed strategies is far richer than the set of behavioral strategies

- sufficiently rich so that every game in extensive form has saddle-point equilibria in mixed policies.

For large classes of games, the set of behavioral policies is already sufficiently rich so that one can already find saddle-point equilibria in behavioral policies.

Moreover, since the number of DOF for behavioral policies is much lower, finding such equilibria is computationally much simpler.

Recursive Computation of Equilibria for Feedback Games

Recursive Computation of Eq. for Feedback Games

Procedure to compute **behavioral** saddle-point equilibria (assume P_1 is **first-acting**), game has K stages.

Step 1. Construct a matrix game corresponding to each IS of P_2 at the Kth stage. Each matrix game will have

- one action of P_1 for each edge entering the information set
- one action of P_2 for each edge leaving the information set.

Step 2. Compute **mixed** saddle-point equilibria for each of the matrix games constructed.

- P_2's behavioral security/saddle-point sub-policy σ_K^* for the Kth stage should map to each IS the probability distribution over actions corresponding to P_2's mixed security policy in the corresponding matrix game.

Recursive Computation of Eq. for Feedback Games

Step 3. Replace each IS for P_2 at the Kth stage by the value of the corresponding mixed saddle-point equilibrium.

- link leading to this value should be labeled with P_1's mixed policy in the corresponding matrix game.

Step 4. The behavioral security/saddle-point sub-policy γ_K^* for P_1 at the Kth stage maps to each IS of P_1 the probability distribution in the link corresponding to the most favorable saddle-point value for P_1.

The policies (γ_K^*, σ_K^*) have the property that, for all policies $\gamma_1, \gamma_2, \ldots, \gamma_{K-1}$ and $\sigma_1, \sigma_2, \ldots, \sigma_{K-1}$ for the stages prior to K,

$$J\left((\{\gamma_{1\ldots K-1}, \boxed{\gamma_K^*}\}, \{\sigma_{1\ldots K-1}, \boxed{\sigma_K^*}\})\right) \leq J\left((\{\gamma_{1\ldots K-1}, \boxed{\gamma_K}\}, \{\sigma_{1\ldots K-1}, \boxed{\sigma_K^*}\})\right), \quad \forall \gamma_K$$

$$J\left((\{\gamma_{1\ldots K-1}, \boxed{\gamma_K^*}\}, \{\sigma_{1\ldots K-1}, \boxed{\sigma_K^*}\})\right) \geq J\left((\{\gamma_{1\ldots K-1}, \boxed{\gamma_K^*}\}, \{\sigma_{1\ldots K-1}, \boxed{\sigma_K}\})\right), \quad \forall \sigma_K$$

as required for a feedback pure saddle-point equilibrium

Recursive Computation of Eq. for Feedback Games

Step 5. Replace each IS for P_1 at the Kth stage by the value corresponding to the link selected by P_1.

At this point we have a game with $K - 1$ stages, and we compute the sub-policies $(\gamma^*_{K-1}, \sigma^*_{K-1})$ using the same procedure described above.

Algorithm is iterated until all sub-policies have been obtained.

At every stage, we guarantee by construction that the conditions for a feedback behavioral saddle-point equilibrium hold.

Mixed vs. Behavioral Order Interchangeability

Mixed vs. Behavioral Order Interchangeability

Lemma 8.2 (Mixed vs. behavioral order interchangeability).

For every feedback game G in extensive form with a finite number of stages, if (γ_m^*, σ_m^*) is a mixed saddle-point equilibrium (SPE) and $(\gamma_\flat^*, \sigma_\flat^*)$ is a behavioral SPE then

1. $(\gamma_m^*, \sigma_\flat^*)$ is a saddle-point equilibrium for a game in which the action space of P_1 is the set of mixed polices and the action space of P_2 is the set of behavioral polices.
2. $(\gamma_\flat^*, \sigma_m^*)$ is a saddle-point equilibrium for a game in which the action space of P_1 is the set of behavioral polices and the action space of P_2 is the set of mixed polices.

And all four equilibria have exactly the same value.

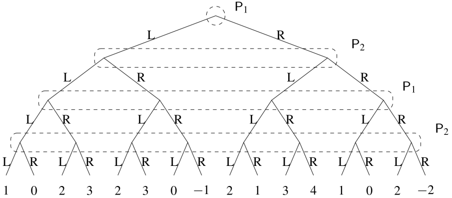
Consequence: little incentive in working in mixed policies since these are computationally more difficult and do not lead to benefits for the players.

Non-Feedback Games

Non-Feedback Games

For non-feedback there is no known general recursive algorithm to compute saddle-point equilibria (pure, behavioral, or mixed).

For non-feedback games, although mixed saddle-point equilibria always exist, behavioral saddle-point equilibria may not exist.



This game does not have behavioral saddle-point equilibria.

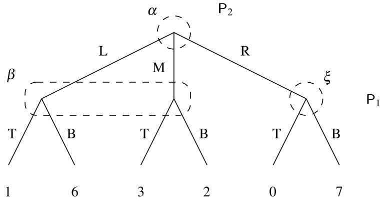
Each player knows nothing other than the current stage.

In fact, both players do not even recall their previous choices.

Practice Exercises

Practice Exercises

8.1. Consider the game in extensive form



Find mixed saddle-point equilibria, using policy domination and the graphical method.

Solution. Enumerate the policies for P_1 and P_2 as follows:

P_1 :

IS	γ_1	γ_2	γ_3	γ_4
β	T	T	B	B
ξ	T	B	T	B

P_2 :

IS	σ_1	σ_2	σ_3
α	L	M	R

Practice Exercises

In this case, we obtain

$$A_{\text{ext}} = \underbrace{\left.\begin{bmatrix} 1 & 3 & 0 \\ 1 & 3 & 7 \\ 6 & 2 & 0 \\ 6 & 2 & 7 \end{bmatrix}\right\}}_{\substack{P_2 \text{ policies} \\ (\sigma_1, \sigma_2, \sigma_3)}} \begin{matrix} P_1 \text{ policies} \\ (\gamma_1, \gamma_2, \gamma_3, \gamma_4) \end{matrix}$$

Using policy domination, matrix is reduced as follows:

$$A_{\text{ext}} = \xrightarrow[\text{row 1 weakly dominates over 2}]{\text{row 3 weakly dominates over 4}} A_{\text{ext}}^{\dagger} := \begin{bmatrix} 1 & 3 & 0 \\ 6 & 2 & 0 \end{bmatrix}$$

$$\xrightarrow[\text{col. 1 strictly dominates over 3}]{} A_{\text{ext}}^{\ddagger} := \begin{bmatrix} 1 & 3 \\ 6 & 2 \end{bmatrix}$$

Practice Exercises

Using the graphical method we obtain

$$V_m(A^{\ddagger}_{\text{ext}}) = \frac{3}{8} \qquad y^* = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} \qquad z^* = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \end{bmatrix}$$

which leads to the value and mixed saddle-point equilibrium for the original game:

$$V_m(A_{\text{ext}}) = \frac{3}{8} \qquad y^* = \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{1}{3} \\ 0 \end{bmatrix} \qquad z^* = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \\ 0 \end{bmatrix}$$

Practice Exercises

8.2. For the game

$$A_{\text{ext}} = \underbrace{\left.\begin{bmatrix} 1 & 3 & 0 \\ 1 & 3 & 7 \\ 6 & 2 & 0 \\ 6 & 2 & 7 \end{bmatrix}\right\}}_{\substack{P_2 \text{ policies} \\ (\sigma_1, \sigma_2, \sigma_3)}} \begin{matrix} P_1 \text{ policies} \\ (\gamma_1, \gamma_2, \gamma_3, \gamma_4) \end{matrix}$$

are there security policies other than these?

$$V_m(A_{\text{ext}}) = \frac{8}{3} \qquad y^* = \begin{bmatrix} \frac{2}{3} \\ 0 \\ 0 \\ \frac{1}{3} \end{bmatrix} \text{ or } \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{1}{3} \\ 0 \end{bmatrix}, \qquad z^* = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{6} \\ 0 \end{bmatrix}$$

Practice Exercises

Solution to Exercise 8.2.

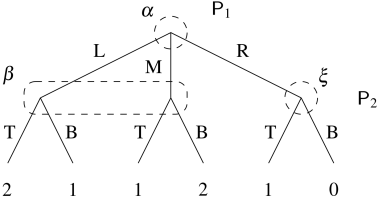
Yes, we can take any convex combination of the policies for P_1

$$y^* = \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{\mu}{3} \\ \frac{1-\mu}{3} \end{bmatrix} \qquad \forall \mu \in [0,1]$$

Moreover, we also lost other equilibria when we used weak domination to get the matrix $A^{\dagger}_{\mathbf{ext}}$, from which this equilibrium was computed.

Practice Exercises

8.3.(a) Compute pure or behavioral saddle-point equilibria for the game in extensive form



Solution. Construction of matrix games for the different IS

$$A_\beta = \underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}_{\substack{P_2 \text{ choices} \\ (T,B)}} \left.\vphantom{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}\right\} \begin{matrix} P_1 \text{ choices} \\ (L, M) \end{matrix} \qquad A_\xi = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\substack{P_2 \text{ choices} \\ (T,B)}} \left.\vphantom{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}\right\} \begin{matrix} P_1 \text{ choice} \\ (R) \end{matrix}$$

Practice Exercises

Solution of matrix games for the different information sets

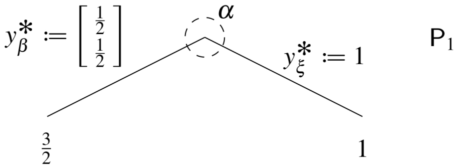
$$V_m(A_\beta) = \frac{3}{2}, \qquad y^*_\beta := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \qquad z^*_\beta := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$V_m(A_\xi) = 1, \qquad y^*_\xi := 1, \qquad z^*_\xi := \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$P_2 \text{ policy } = \begin{cases} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} & \text{if IS} = \beta \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if IS} = \xi \end{cases}$$

Practice Exercises

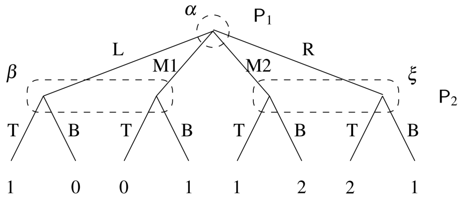
Replacement of information sets by the values of the game.



$$P_1 \text{ policy } = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{if IS} = \alpha \right.$$

Practice Exercises

8.3.(b) Compute pure or behavioral saddle-point equilibria for the game in extensive form



Solution. Construction of matrix games for the different IS

$$A_\beta = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\substack{P_2 \text{ choices} \\ (T,B)}} \Big\} \; \begin{matrix} P_1 \text{ choices} \\ (L, M1) \end{matrix} \qquad A_\xi = \underbrace{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}}_{\substack{P_2 \text{ choices} \\ (T,B)}} \Big\} \; \begin{matrix} P_1 \text{ choices} \\ (M2, R) \end{matrix}$$

Practice Exercises

Solution of matrix games for the different information sets

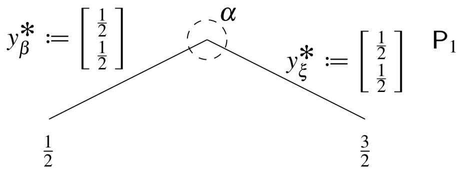
$$V_m(A_\beta) = \frac{1}{2}, \qquad y^*_\beta := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \qquad z^*_\beta := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$V_m(A_\xi) = \frac{3}{2}, \qquad y^*_\xi := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \qquad z^*_\xi := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$P_2 \text{ policy } = \begin{cases} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} & \text{if IS} = \beta \text{ or } \xi \end{cases}$$

Practice Exercises

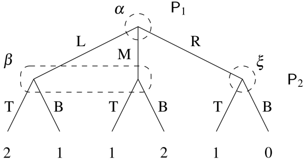
Replacement of information sets by the values of the game.

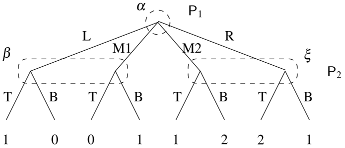


$$P_1 \text{ policy } = \begin{cases} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{bmatrix} & \text{if IS} = \alpha \end{cases}$$

Practice Exercises

8.4. Consider the games in extensive form





and behavioral policies of the form

P_1 :

IS	1	2	$\cdots$	m
α	y_1^α	y_2^α	$\cdots$	y_m^α

$$y_i^\alpha \geq 0, \quad \forall i, \quad \sum_i y_i^\alpha = 1$$

P_2 :

IS	1	2
β	z_1^β	z_2^β
ξ	z_1^ξ	z_2^ξ

$$z_1^\beta, z_2^\beta \geq 0, \quad z_1^\beta + z_2^\beta = 1$$
$$z_1^\xi, z_2^\xi \geq 0, \quad z_1^\xi + z_2^\xi = 1$$

Practice Exercises

1. Show the expected cost in games can be written in the form

$$J(y^\alpha, z^\beta, z^\xi) := y^{\alpha\prime} A^\beta z^\beta + y^{\alpha\prime} A^\xi z^\xi$$

where

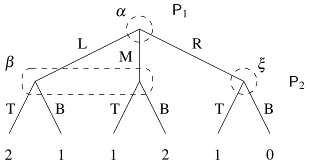
$$y^\alpha := [y_1^\alpha \quad y_2^\alpha \quad \cdots \quad y_m^\alpha]', \qquad z^\beta := \begin{bmatrix} z_1^\beta & z_2^\beta \end{bmatrix}' \qquad z^\xi := \begin{bmatrix} z_1^\xi & z_2^\xi \end{bmatrix}'$$

2. Formulate the computation of the security policy for P_2 as a linear program. The matrices A^β and A^ξ in the eq. $J(y^\alpha, z^\beta, z^\xi)$ should appear in this linear program.
3. Formulate the computation of the security policy for P_1 as a linear program. The matrices A^β and A^ξ in the eq. $J(y^\alpha, z^\beta, z^\xi)$ should also appear in this linear program.
4. Use the two linear programs above to find behavioral saddle-point equilibria for the games in the Figure numerically using MATLAB®

Practice Exercises

Solution to Exercise 8.4.

1. For the game



the expected cost is given

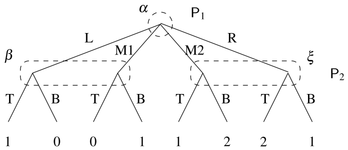
$$J(y^\alpha, z^\beta, z^\xi) = 2 \times y_1^\alpha z_1^\beta + 1 \times y_1^\alpha z_2^\beta + 1 \times y_2^\alpha z_1^\beta + 2 \times y_2^\alpha z_2^\beta + 1 \times y_3^\alpha z_1^\xi + 0 \times y_3^\alpha z_2^\xi$$

$$= y^{\alpha\prime} \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0 & 0 \end{bmatrix} z^\beta + y^{\alpha\prime} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} z^\xi$$

Practice Exercises

Solution to Exercise 8.4.

1. And for the game



the expected cost is given

$$J(y^\alpha, z^\beta, z^\xi) = 1 \times y_1^\alpha z_1^\beta + 0 \times y_1^\alpha z_2^\beta + 0 \times y_2^\alpha z_1^\beta + 1 \times y_2^\alpha z_2^\beta + 1 \times y_3^\alpha z_1^\xi + 2 \times y_3^\alpha z_2^\xi + 2 \times y_4^\alpha z_1^\xi + 1 \times y_4^\alpha z_2^\xi$$

$$= y^{\alpha\prime} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} z^\beta + y^{\alpha\prime} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} z^\xi$$

Practice Exercises

2. The security policy for P_2 arises out of the following minimax computation:

$$\underline{V}(G) := \max_{z^\beta, z^\xi \in \mathcal{Z} \subset \mathbb{R}^2} \min_{y^\alpha \in \mathcal{Y} \subset \mathbb{R}^3} y^{\alpha\prime} A^\beta z^\beta + y^{\alpha\prime} A^\xi z^\xi$$

Focusing on the inner minimization, we conclude that

$$\min_{y^\alpha \in \mathcal{Y}} y^{\alpha\prime} A^\beta z^\beta + y^{\alpha\prime} A^\xi z^\xi = \min_{i \in \{1,2,\ldots,m\}} \underbrace{(A^\beta z^\beta + A^\xi z^\xi)_i}_{i\text{th entry of } A^\beta z^\beta + A^\xi z^\xi}$$

$$= \max \left\{ v : v \leq (A^\beta z^\beta + A^\xi z^\xi)_i, \quad \forall i \right\}$$

Practice Exercises

Therefore

$$\underline{V}(G) = \max_{z^\beta, z^\xi \in \mathcal{Z}} \max \left\{ v : v \leq (A^\beta z^\beta + A^\xi z^\xi)_i, \quad \forall i \right\}$$

$$\underbrace{\begin{aligned} &= \text{maximum } v \\ &\quad \text{subject to} \quad A^\beta z^\beta + A^\xi z^\xi \quad \dot{\geq} v\mathbf{1} \\ &\qquad \left.\begin{aligned} z^\beta &\dot{\geq} 0 \\ \mathbf{1} z^\beta &= 1 \end{aligned}\right\} \quad (z^\beta \in \mathcal{Z}) \\ &\qquad \left.\begin{aligned} z^\xi &\dot{\geq} 0 \\ \mathbf{1} z^\xi &= 1 \end{aligned}\right\} \quad (z^\xi \in \mathcal{Z}) \end{aligned}}_{\text{optimization over 4+1 parameters } (v, z_1^\beta, z_2^\beta, z_1^\xi, z_2^\xi)}$$

Practice Exercises

3. The security policy for P_1 arises out of the following minimax computation:

$$\bar{V}(G) := \min_{y^\alpha \in \mathcal{Y} \subset \mathbb{R}^3} \max_{z^\beta, z^\xi \in \mathcal{Z} \subset \mathbb{R}^2} y^{\alpha\prime} A^\beta z^\beta + y^{\alpha\prime} A^\xi z^\xi$$

$y^{\alpha\prime} A^\beta z^\beta$ only depends on z^β and $y^{\alpha\prime} A^\xi z^\xi$ only on z^ξ, then

$$\begin{aligned}
\max_{z^\beta, z^\xi \in \mathcal{Z}} y^{\alpha\prime} A^\beta z^\beta + y^{\alpha\prime} A^\xi z^\xi &= \max_{z^\beta \in \mathcal{Z}} y^{\alpha\prime} A^\beta z^\beta + \max_{z^\xi \in \mathcal{Z}} y^{\alpha\prime} A^\xi z^\xi \\
&= \max_{j \in \{1,2\}} \underbrace{(y^{\alpha\prime} A^\beta)_j}_{j\text{th entry of } y^{\alpha\prime} A^\beta} + \max_{j \in \{1,2\}} \underbrace{(y^{\alpha\prime} A^\xi)_j}_{j\text{th entry of } y^{\alpha\prime} A^\xi} \\
&= \min \left\{ v_1 : v_1 \geq (y^{\alpha\prime} A^\beta)_j, \quad \forall j \right\} + \min \left\{ v_2 : v_2 \geq (y^{\alpha\prime} A^\xi)_j, \quad \forall j \right\} \\
&= \min \left\{ v_1 + v_2 : v_1 \geq (y^{\alpha\prime} A^\beta)_j, \quad \forall j, \quad v_2 \geq (y^{\alpha\prime} A^\xi)_j, \quad \forall j \right\}
\end{aligned}$$

Practice Exercises

Therefore

$$\bar{V}(G) = \min_{y^\alpha \in \mathcal{Y}} \min \left\{ v_1 + v_2 : v_1 \geq (y^{\alpha\prime} A^\beta)_j, \quad \forall j, \quad v_2 \geq (y^{\alpha\prime} A^\xi)_j, \quad \forall j \right\}$$

$$= \text{minimum } v_1 + v_2$$

$$\underbrace{\begin{array}{lll} \text{subject to} & A^{\beta\prime} y^\alpha & \dot{\leq} v_1 \mathbf{1} \\ & A^{\xi\prime} y^\alpha & \dot{\leq} v_2 \mathbf{1} \\ & \left.\begin{array}{l} y^\alpha \dot{\geq} 0 \\ \mathbf{1} y^\alpha = 1 \end{array}\right\} & (y \in \mathcal{Y}) \end{array}}_{\text{optimization over } m+2 \text{ parameters } (v_1, v_2, y_1^\alpha, y_2^\alpha, \ldots, y_m^\alpha)}$$

Practice Exercises

4. CVX code to compute the behavioral SPE for the first game

```
Abeta = [2,1; 1,2; 0,0];
Axi =[0,0; 0,0; 1,0];

% P1 security policy
 cvx_begin
  variables v zbeta(size(Abeta,2)) zxi(size(Axi,2))
  maximize v
  subject to
   Abeta*zbeta + Axi*zxi >= v
   zbeta >= 0
   sum(zbeta)==1
   zxi >= 0
   sum (zxi)==1
cvx_end
```

```
% P2 security policy
cvx_begin
 variables v1 v2 yalpha(size(Abeta,1))
 minimize v1 + v2
 subject to
  Abeta'*yalpha <= v1
  Axi'*yalpha <= v2
  yalpha >= 0
  sum(yalpha)==1
cvx_end
```

which results in the following values

```
v = 1.0000
zbeta = 0.5000   0.5000
zxi = 1.0000   0.0000
v1 = 1.4084 e-09
v2 = 1.0000
yalpha = 0.0000   0.0000   0.0000
```

Practice Exercises

4. CVX code to compute the behavioral SPE for the second game

```
Abeta = [1,0; 0,1; 0,0; 0,0];
Axi =[0,0; 0,0; 1,2; 2,1];
 ...
```

which results in the following values

```
v = 0.5000
zbeta = 0.5000   0.5000
zxi = 0.5000   0.5000
v1 = 0.5000
v2 = 2.9174 e-09
yalpha = 0.5000   0.5000   0.0000   0.0000
```

End of Lecture

08 - Stochastic Policies for Games in Extensive Form

Questions?