CS272 - Theoretical Foundations of Learning, Decisions, and Games

Lecture 12: General-Sum Games and Nash Equilibria

October 07, 2025

Lecturer: Nika Haghtalab *Readings: None*
Scribe: Karim Abdel Sadek

1 Introduction and brief recap

Last lecture we introduced zero-sum games, where the sum of payoffs of each player is indeed zero. This implies that in such games, the optimal strategic behaviour is to simply win against your opponents. Today we will introduce general sum games, where players may both benefit from playing certain actions, capturing a wider range of situations. For example, consider a situation where you have a buyer and a seller. The former gets utility from the item received, and the latter from the price it is sold at. In this case, it can be the case that both players can manage to get a deal they are both happy with. In general sum games, a notion of optimal play can be more nuanced than in the case of zero-sum ones.

Notation Let's now recall some of the notation we will use throughout the lecture. We have n players, and refer to them by an index $i \in [n]$. The action space is defined by $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$, where A_i is the action space for player i. A *pure action profile* is defined as $\boldsymbol{a} = (a_1, \ldots, a_n)$ where $a_i \in \mathcal{A}_i$. In general, we can write an action profile as $\boldsymbol{a} = (a_i, \boldsymbol{a}_{-i})$ where $\boldsymbol{a}_{-i}$ represents the actions of all players but the i-th one. A mixed strategy for player i is defined as $p_i \in \Delta(\mathcal{A}_i)$, and $\boldsymbol{p} = (p_1, \ldots, p_n)$ denotes a *mixed strategy profile*. A utility function for player i is defined as $U_i : \mathcal{A} \to [0, 1]$. Given a mixed strategy profile, the expected utility is simply $U_i(\boldsymbol{p}) = \mathbb{E}_{a_i \sim p_i}[U_i(a_1, \ldots, a_n)]$.

Zero-sum games and the minimax theorem Last lecture, we proved the minimax theorem, which says that in a two-player zero sum game, with utilities U and $-U$ for the first and second player respectively, we have

$$\max_{\boldsymbol{p} \in \Delta(\mathcal{A}_1)} \min_{\boldsymbol{q} \in \Delta(\mathcal{A}_2)} U(\boldsymbol{p}, \boldsymbol{q}) = \min_{\boldsymbol{q} \in \Delta(\mathcal{A}_2)} \max_{\boldsymbol{p} \in \Delta(\mathcal{A}_1)} U(\boldsymbol{p}, \boldsymbol{q}). \tag{1}$$

The quantity in eq. (1) is referred to as the *value of the game*, which we can denote as $\mathrm{value}(U)$. The theorem thus implies that there exists a strategy one could play that guarantees the value of the game, independently of the strategy of the opponent. More precisely, we have the following definitions:

- $p^* \in \Delta(A_1)$ is maximin optimal if $\forall q \in \Delta(A_2)$ we have $U(p^*, q) \geq \mathrm{value}(U)$.
- $q^* \in \Delta(A_2)$ is minimax optimal if $\forall p \in \Delta(A_1)$ we have $U(p, q^*) \leq \mathrm{value}(U)$.

Any pair of maximin/minimax strategies are indeed what we expect optimal agents to play in a zero-sum game. In fact, much of the progress for superhuman AI in games such as Go, chess or 2-player poker indeed comes from designing systems that can compute and play such policies [Silver et al., 2016, 2017, Brown and Sandholm, 2019].

2 General sum games

In general-sum games, we often don't talk about optimality of play, instead we describe strategy profiles where players have no "reason" to change their (possibly mixed) strategy. Such notion of stability is broadly referred to as as an *equilibrium.*. Many notions for defining such equilibria exist, which we will see throughout the course.

To make this more concrete, let's consider the following normal form game, also known as Prisoner's dilemma, that captures the following strategies setting: Two members of a criminal gang are arrested and placed in separate cells, with no means of communication. The police admit they lack sufficient evidence to convict either for a serious 3-year sentence. If both prisoners remain silent, the police can only convict them on a lesser charge, leading to a 1-year sentence each.

However, the police offer each prisoner a deal independently: If one confesses and provides evidence against the other, the confessor goes free (0 years), while the other receives the full 3-year sentence. If both confess, their cooperation provides enough evidence to convict both, but the police reward their cooperation by reducing each sentence to 2 years.

	C	NC
C	-2, -2	0, -3
NC	-3, 0	-1, -1

Formally, each player has two possible actions: either confess (C) or non-confess (NC) that they committed a crime and the payoffs are as described above. Note that, given any possible action of the column player, the row player will increase her own payoff by confessing. Such and action is called a *dominant strategy*. So, two rational players will end up both confessing and getting a 2-year sentence. This is called a dominant strategy equilibrium. Formally:

Definition 2.1 (Dominant Strategy)**.** A pure strategy a_i is a dominant strategy (DS) for player i if the following holds:

$$\forall \boldsymbol{a}_{-i} \in \mathcal{A}_{-i},\ \forall a_i' \in \mathcal{A}_i \quad U_i(a_i, \boldsymbol{a}_{-i}) \geq U_i(a_i', \boldsymbol{a}_{-i})$$

Intuitively, a dominant strategy is an action that is always optimal to play, where optimality means that given any other action of our opponent, playing the dominant one gives higher utility. We can now define the notion of a *dominant strategy equilibrium* (DSE).

Definition 2.2 (Dominant Strategy Equilibrium)**.** A pure action profile $\boldsymbol{a} = (a_1, \ldots, a_n)$ is a DSE if for all players $i \in [n]$, we have that a_i is a dominant strategy.

A DSE is simply an action profile where every player is playing a dominant strategy. Note that the action profile (C, C) was indeed a DSE in the Prisoner's dilemma. While playing a dominant strategy seems to capture a good notion of optimal play, such is not guaranteed to exist. For example, in Rock-Paper-Scissors, a dominant strategy does not exist, as what I should play depends on the strategy of my opponent. We will now relax this notion of stability by introducing the Nash Equilibrium (NE) [Nash].

Definition 2.3 (Nash Equilibrium)**.** A mixed strategy profile $\boldsymbol{p}^* = (p_1^*, \ldots, p_n^*)$ is a mixed strategy Nash equilibrium (MNE) if and only if

$$\forall i \in [n],\ \forall p_i' \in \Delta(\mathcal{A}_i), \quad U_i(p_i^*, \boldsymbol{p}_{-i}^*) \geq U_i(p_i', \boldsymbol{p}_{-i}^*)$$

A Nash Equilibrium is a pure NE (PNE) if all $p_1^*, \ldots, p_n^*$ are pure strategies.[1]

A NE is a strategy profile where none of the players is incentivized to *unilaterally* deviate and choose a different action. Note that this does not mean that a better outcome for any players does not exist, but only that this cannot be achieved by them changing their action if all the opponents' actions remain fixed.

Observation: Every DSE is a Nash equilibrium, but not vice versa. DSE requires optimality *against all* $\boldsymbol{a}_{-i}$, while Nash only requires optimality *at the currently realized profile* $\boldsymbol{p}^*_{-i}$ (or $\boldsymbol{a}^*_{-i}$).

[1]Note that if deviations to a mixed strategy p_i' is beneficial to player i, there is also a pure strategy to which the player prefers to deviate. So, its also sufficient to consider just deviations to pure strategies.

MNE and stability As a small digression, we can have a closer look at the relation between MNE and some more intuitive stability definition. Let's look at the gradients of the utility function with respect to each strategy p_i, i.e. $\nabla_{p_i} U_i(p^*)$. The question we may have is that if this gradient being equal to zero says something about the MNE. In other words, are these gradients being zero a necessary and sufficient condition for the strategy profile to be an MNE?

- *"If p^* is MNE $\not\Longrightarrow$ $\forall i \in [n],\ \nabla_{p_i} U_i(p^*) = 0$"* Intuitively, this is because we could be 'on the wall' of the simplex. In other words, there could exist directions under which p_i could be changed to obtain higher utility (meaning the gradient is non-zero), but those may not be valid probability distributions. The action profile (C, C) in the Prisoner's Dilemma, would be an example of this being the case.
- *"If $\forall i \in [n],\ \nabla_{p_i} U_i(p^*) = 0 \implies p^*$ an MNE."* Given the fact that U_i linear in p_i, this follows. Our game has a bilinear shape, and all the stable points are well-behaved. In other words, once we fix the strategy of the opponents, we have a convex (or concave) structure, which allows the above implication to follow.

3 On the existence of Mixed Nash Equilibria in finite games

We are now going to prove one of the most fundamental results in game theory: given a game with finite players, action space, and range of utilities, a mixed Nash Equilibrium *always* exists.

Theorem 3.1 (Existence of a Nash Equilibrium)**.** *Every game with a finite number of players n, a finite action space $\mathcal{A}$, and utilities within a finite range, has a Nash Equilibrium in mixed strategies.*

For the proof of the theorem, we will rely on Brouwer's Fixed-Point theorem.

Theorem 3.2 (Brouwer's Fixed-Point)**.** *Let $\mathcal{X}$ be a nonempty, compact, convex subset of $\mathbb{R}^d$ and $f : \mathcal{X} \to \mathcal{X}$ a continuous function. Then,*

$$\exists x^* \in \mathcal{X} \quad \textit{such that} \quad f(x^*) = x^*,$$

and x^* is called a fixed point.

The idea of the proof for theorem 3.1 is to apply Brouwer's fixed point to our case. Suppose we define a function $f : \Delta(\mathcal{A}) \to \Delta(\mathcal{A})$ that captures what the best response for each player would be. More formally, let $f_i(\boldsymbol{p}) = \arg\max_{p_i'} U_i(p_i', \boldsymbol{p}_{-i})$. If we can show that for all players i we have that $f_i(p_i^*, \ldots, p_n^*) = p_i^*$, i.e., that $\boldsymbol{p}^*$ is a fixed point, then it must be the case that such strategy profile is a Nash equilibrium. However, note that in the current form, f is not necessarily continuous (given the argmax), and thus we must find a different form or regularize it so we can apply the fixed-point theorem.

Let's now do so and prove our theorem below.

Proof of theorem 3.1.

Let $\Delta(A) = \prod_{i=1}^{n} \Delta(A_i)$ be the product of the players' mixed-strategy simplices; $\Delta(A)$ is nonempty, compact, and convex. Consider a function $\Phi : \Delta \to \Delta$. and for $\boldsymbol{p} = (p_1, \dots, p_n) \in \Delta(A)$ define, for each player i,

$$\Phi_i(\boldsymbol{p}) \;:=\; \operatorname*{arg\,max}_{\rho_i \in \Delta(A_i)} \Big\{ U_i(\rho_i, \boldsymbol{p}_{-i}) \;-\; \|\rho_i - p_i\|_2^2 \Big\}, \qquad \Phi(\boldsymbol{p}) := (\Phi_1(\boldsymbol{p}), \dots, \Phi_n(\boldsymbol{p})).$$

Claim 1: Φ is a continuous map. To see this, we can apply Berge's theorem [Berge, 1963], also known as the maximum-theorem argument. This result states that given a continuous objective with a unique maximizer, the entire map is continuous. It is clearly the case that our objective is continuous in p_i, so we are left to show that the maximizer is unique. To see this, note that $\|\rho_i - p_i\|_2^2$ is strictly convex. Thus, when we consider the entire objective $U_i(\rho_i, \boldsymbol{p}_{-i}) - \|\rho_i - p_i\|_2^2$, that is strictly concave (since we are subtracting from a linear function a strictly convex one), which implies having a unique maximizer as desired. Thus, by Berge's theorem, it follows that $\Phi : \Delta(A) \to \Delta(A)$ is continuous as desired.

By Brouwer's Fixed-Point theorem, it now follows that there exists $\boldsymbol{p}^* \in \Delta(A)$ with $\Phi(\boldsymbol{p}^*) = \boldsymbol{p}^*$. Now we are only left to show that $\boldsymbol{p}^*$ is a mixed Nash equilibrium.

Let's suppose by contradiction that this is not the case. This means that for some player i, there exists $\rho_i \in \Delta(A_i)$ such that

$$\delta \;:=\; U_i(\rho_i, \boldsymbol{p}^*_{-i}) - U_i(p_i^*, \boldsymbol{p}^*_{-i}) \;>\; 0.$$

The idea is now to construct a $\hat{\rho}$ that is better than $\boldsymbol{p}^*$ under the objective optimized by Φ, given the above inequality being true. That would contradict $\boldsymbol{p}^*$ being a fixed point.

More precisely, take $\hat{\rho}_i = (1 - \alpha)p_i^* + \alpha\rho_i$ for a value of $\alpha \in (0, 1]$ that we will determine later. By rearranging, we can rewrite the above equation as $\hat{\rho}_i - p_i^* = \alpha(\rho_i - p_i^*)$. Now, given $\boldsymbol{p}^* = (p_1^*, \dots, p_n^*)$ define $F_i(x) = U_i(x, \boldsymbol{p}^*_{-i}) - \|x - p_i^*\|_2^2$.

Final Claim: It is sufficient now to show that we can get a contradiction by proving that $F_i(\hat{\rho}_i) > F_i(p_i^*)$. This is because in that case, it means that $p_i^* \neq \arg\max_x F_i(x)$, i.e. that p^* is not a unique maximizer of the regularized objective we introduced.

To see this, set α as follows

$$\alpha \ := \ \min\Big\{1,\ \frac{\delta}{\|\rho_i - p_i^*\|_2^2}\Big\}.$$

We now make the following observation (Obs 1): $\|\hat{\rho}_i - p_i^*\|_2^2 = \big\|\alpha(\rho_i - p_i^*)\big\|_2^2 = \alpha^2\,\|\rho_i - p_i^*\|_2^2$.

Given the linearity of the utility, we also have the following (Obs 2):

$$\begin{aligned}
U_i(\hat{\rho}_i, \boldsymbol{p}_{-i}^*) - U_i(p_i^*, \boldsymbol{p}_{-i}^*) &= U_i\big((1-\alpha)p_i^* + \alpha\rho_i,\ \boldsymbol{p}_{-i}^*\big) - U_i(p_i^*, \boldsymbol{p}_{-i}^*) \\
&= (1-\alpha)\,U_i(p_i^*, \boldsymbol{p}_{-i}^*) + \alpha\,U_i(\rho_i, \boldsymbol{p}_{-i}^*) - U_i(p_i^*, \boldsymbol{p}_{-i}^*) \\
&= \alpha\Big(U_i(\rho_i, \boldsymbol{p}_{-i}^*) - U_i(p_i^*, \boldsymbol{p}_{-i}^*)\Big) \\
&= \alpha\,\delta.
\end{aligned}$$

Now, combining Obs 1 and Obs 2, what we wanted to show follows. Take the difference between the two F_i, and we have

$$\begin{aligned}
F_i(\hat{\rho}_i; \boldsymbol{p}_{-i}^*) - F_i(p_i^*; \boldsymbol{p}_{-i}^*) &= \Big[U_i(\hat{\rho}_i, \boldsymbol{p}_{-i}^*) - \|\hat{\rho}_i - p_i^*\|_2^2\Big] \ - \ \Big[U_i(p_i^*, \boldsymbol{p}_{-i}^*) - \|p_i^* - p_i^*\|_2^2\Big] \\
&= \Big(U_i(\hat{\rho}_i, \boldsymbol{p}_{-i}^*) - U_i(p_i^*, \boldsymbol{p}_{-i}^*)\Big) \ - \ \|\hat{\rho}_i - p_i^*\|_2^2 \\
&= \alpha\,\delta \ - \ \alpha^2\,\|\rho_i - p_i^*\|_2^2.
\end{aligned}$$

Given our choice of α, it follows that $\alpha\,\delta \ - \ \alpha^2\,\|\rho_i - p_i^*\|_2^2 > 0$. This contradicts that $p_i^* = \Phi_i(p^*)$ uniquely maximizes the regularized objective. Hence we reached a contradiction, and p^* is a mixed Nash equilibrium, concluding the proof.

$\square$

References

Claude Berge. *Topological Spaces*. Oliver and Boyd, London, 1963.

Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456): 885–890, 2019.

Jr. Nash, John F. Equilibrium points in N-person games. 36(1):48–49. ISSN 0027-8424. doi: 10.1073/pnas.36.1.48. URL `https://doi.org/10.1073/pnas.36.1.48`.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*, 2017.