

Modern Algorithmic Game Theory

Martin Schmid

Department of Applied Mathematics
Charles University in Prague

October 15, 2025

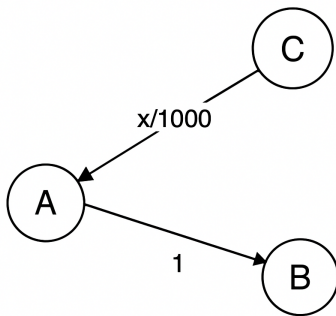
Recap

- In the last two weeks, we started talking about normal-form games, introduced two solution concepts and their properties, and finally described Support Enumeration, an algorithm that finds all possible Nash equilibria in a given game.
- The problematic part about this algorithm is that its runtime is exponential in the number of actions and solving larger games quickly becomes intractable.
- Today, we will first focus on a couple of multi-agent dynamics in non-zero-sum games
- We will then come back to zero-sum games and introduce the first self-play style algorithm – *Fictitious Play*

An aerial photograph of a frozen river or lake. The ice is a mix of white and light blue, with dark, winding channels of water or meltwater. A large, dark, textured island is situated in the upper center. A semi-transparent white rectangular box is centered over the image, containing the title text.

Unexpected Dynamics in Game Theory

Congestion Games and Price of Anarchy



Congestion Games and Price of Anarchy

First, let us compute the minimum possible delay. Suppose x drivers go to road 2 and $1000 - x$ go to road 1. Then, the total delay is

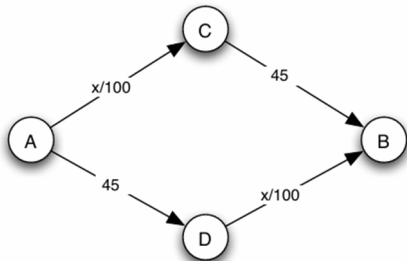
$$\frac{x^2}{1000} + (1000 - x).$$

This expression is minimized when $x \approx 500$, that is, 500 drivers go to road 2 and the other 500 to road 1. The total delay is then

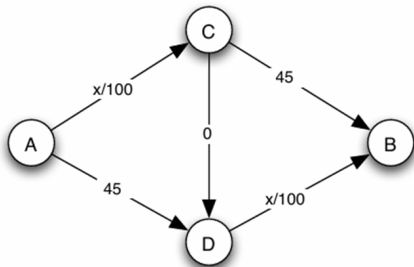
$$500 \times \frac{1}{2} + 500 \times 1 \approx 750 \text{ minutes.}$$

Braess's Paradox

- 4,000 agents



(a) Original roads/actions

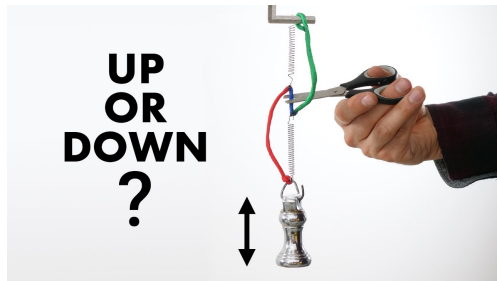


(b) Adding a new road/action

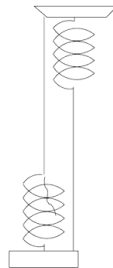
Braess's Paradox

- This actually sometimes happens!
- Could this happen in single agent game?
- Are agents rational?

Braess's Paradox in Springs



(a) Before



(b) After

Figure 2: Strings and springs. Severing a taut string lifts a heavy weight.

Tragedy of the Commons

- Shared resource
- Getting more increases your reward, but puts strain on the shared resource



Tragedy of the Commons

Commonize Costs–Privatize Profits Game

	extra	preserve
extract	(50, 50)	(80, 20)
preserve	(20, 80)	(100, 100)

Table: CC–PP Game

An aerial photograph of a wide, frozen river. The river is mostly covered in a light blue-grey ice, with darker, more textured areas of open water or thinner ice interspersed. A large, irregularly shaped island or peninsula is located in the upper center of the frame. A prominent, dark, winding line, possibly a crack or a narrow channel, runs diagonally across the right side of the image. The overall scene is desolate and cold.

Evaluating the Quality of Strategy Profiles

Motivation

How can we evaluate the quality of an agent, i.e. how well it plays a game?

- In single-agent environments we can easily determine how good an agent is based on the score it achieves (e.g. in Atari)
- However, the agent's performance in multi-agent environments directly depends on the quality of its opponents
- ELO-based comparisons are problematic due to intransitivity
- We will use measures that tell us how “close” to an optimal policy we are in terms of performance rather than distance (e.g. KL divergence)

Metrics

We define the following metrics:

- the incentive of player i to deviate

$$\delta_i(\pi) = \max_{\pi'_i} u_i(\pi'_i, \pi_{-i}) - u_i(\pi) = u_i(b(\pi_{-i}), \pi_{-i}) - u_i(\pi)$$

- the total incentive to deviate

$$NashConv(\pi) = \sum_{i \in \mathcal{N}} \delta_i(\pi)$$

- the average incentive to deviate

$$Exploitability(\pi) = \frac{NashConv(\pi)}{|\mathcal{N}|}$$

Metrics in Zero-Sum Games

- In zero-sum games, the definition of $NashConv(\pi)$ simplifies to

$$NashConv(\pi) = \sum_{i \in \mathcal{N}} u_i(b(\pi_{-i}, \pi_{-i})),$$

- In two-player zero-sum games ($u = u_1 = -u_2$), $NashConv(\pi)$ can be written as

$$NashConv(\pi) = \max_{\pi'_i} u(\pi'_i, \pi_{-i}) - \min_{\pi'_{-i}} u(\pi_i, \pi'_{-i})$$

ϵ -Nash Equilibrium

Definition: ϵ -Nash Equilibrium

Strategy profile (π_i, π_{-i}) is an ϵ -**Nash equilibrium** if none of the players can improve by more than ϵ by unilaterally deviating from their policy. Mathematically,

$$\forall i \in \mathcal{N}, \forall \pi'_i \in \Pi_i : u_i(\pi_i, \pi_{-i}) \geq u_i(\pi'_i, \pi_{-i}) - \epsilon$$

- Iterative algorithms may not be able to find an exact Nash equilibrium due to finite computation time and/or numerical instabilities
- Therefore, we need to use a looser definition of an optimal strategy profile
- If a strategy profile π is an ϵ -Nash equilibrium, it holds that $\max_{i \in \mathcal{N}} \delta_i(\pi) \leq \epsilon$
- Consequently, a strategy profile is a Nash equilibrium $\Leftrightarrow \text{Exploitability}(\pi) = 0$

An aerial photograph of a frozen river or stream. The ice is white and textured, with dark, winding channels of water or meltwater visible. A central text box is overlaid on the image.

Fictitious Play

Fictitious Play

- An iterative algorithm where players repeatedly play against each other and keep track of the **empirical distribution** over their opponent's previously played actions
- Both players **simultaneously** compute a **pure** best response to maximize their expected payoff against their opponent's observed **average** strategy
- The sequence of **average** strategies produced by the algorithm converges in certain classes of games to Nash equilibria; a property called **average-iterate** convergence
- The actual sequence of best-response strategies **does not** converge in general
- Mathematically, a single iteration of the algorithm can be expressed as

$$\bar{\pi}_i^{t+1} \in \left(1 - \frac{1}{t+1}\right) \bar{\pi}_i^t + \frac{1}{t+1} b(\bar{\pi}_{-i}^t),$$

where $\bar{\pi}_i^t = \frac{1}{t} \sum_{k=1}^t \pi_i^k$ is player i 's average strategy and $\bar{\pi}_{-i}^t$ is defined analogously.

Fictitious Play

- Fictitious Play is a belief-based learning process in repeated games
- Each player assumes that the opponent's play is stationary and equal to the empirical distribution of their past actions
- At every round, player i plays a pure best response to that belief
- It is therefore **myopic**: players maximize their current expected payoff given their current belief, without anticipating future consequences or learning the opponent's update rule
- It is not deterministic if multiple best responses exist; the theory usually assumes any of them can be chosen.

Fictitious Play Example

Let's consider the Matching Pennies game and simulate a couple of steps of Fictitious play. Let n_i^t be the running count of the number of times player i played each action.

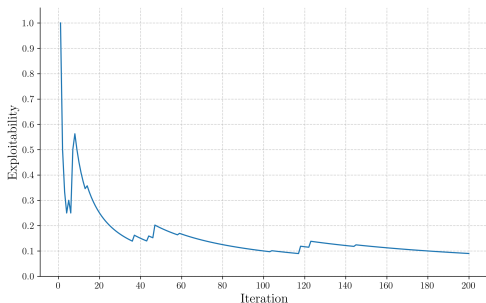
	heads	tails
heads	(1, -1)	(-1, 1)
tails	(-1, 1)	(1, -1)

Table: Matching Pennies

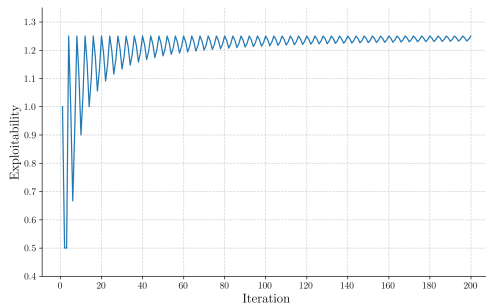
Time	n_1^t	n_2^t	Play
0	(0, 0)	(0, 0)	(h, h)
1	(1, 0)	(1, 0)	(h, t)
2	(2, 0)	(1, 1)	(h, t)
3	(3, 0)	(1, 2)	(t, t)
4	(3, 1)	(1, 3)	(t, t)
5	(3, 2)	(1, 4)	(t, t)
6	(3, 3)	(1, 5)	(t, h)
7	(3, 4)	(2, 5)	(t, h)
8

Average vs. Current Strategy Convergence

- When we best-respond to opponent's previously played action, instead of their average strategy, Fictitious play may not converge to a Nash equilibrium



(a) Best response against the average strategy



(b) Best response against the current strategy

Convergence of Fictitious Play to Pure Strategies

- Let $\{\pi^t\}$ be a sequence of strategy profiles generated by Fictitious Play
- Let us now study the asymptotic behavior of the sequence $\{\pi^t\}$, i.e. the convergence properties of the sequence $\{\pi^t\}$ as $t \rightarrow \infty$
- We say the sequence $\{\pi^t\}$ converges to π^* , if there exists T such that $\forall t \geq T: \pi^t = \pi^*$
- The next proposition formalizes the property that if the FP sequence converges, then it must converge to a Nash equilibrium of the game

Theorem

Let $\{\pi^t\}$ be a sequence of strategy profiles generated by Fictitious Play. If $\{\pi^t\}$ converges to some π^* , then π^* is a pure strategy Nash equilibrium. Furthermore, suppose that for some t , $\pi^t = \pi^*$, where π^* is a **strict** Nash equilibrium. Then $\pi^{t'} = \pi^*$ for all $t' \geq t$.

Convergence of Fictitious Play to Mixed Strategy

Definition

The sequence $\{\pi^t\}$ converges to a mixed strategy profile π^* in the time-average sense, if for each player i and for all actions $a_i \in \mathcal{A}_i$, we have:

$$\lim_{T \rightarrow \infty} \frac{\sum_t I(\pi_i^t = a_i)}{T} = \pi^*(a_i)$$

Theorem

Suppose that the sequence π^t converges to π^* in the time-average sense. Then π^* is a Nash equilibrium.

Convergence of Fictitious Play

Games in which Fictitious Play converges are said to have **the fictitious-play property**. The algorithm has been proven to converge for the following classes of games:

- two-player zero-sum games
- two-player non-zero-sum game, where each player has at most two strategies
- games solvable by Iterated removal of strictly dominated strategies
- identical interest games; games where all players have the same payoff function

Convergence of Fictitious Play

On the other hand, in games such as the **Shapley game**, Fictitious Play can cycle indefinitely and fail to converge, depending on the initial conditions.

	x	y	z
a	(2, 1)	(0, 0)	(1, 2)
b	(1, 2)	(2, 1)	(0, 0)
c	(0, 0)	(1, 2)	(2, 1)

Table: The Shapley Game

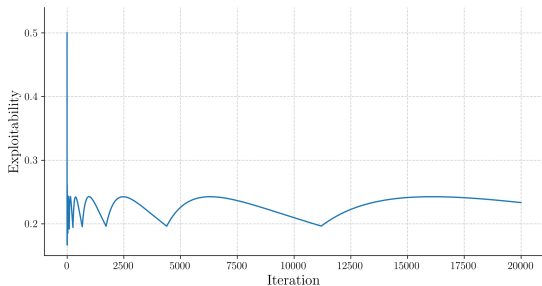


Figure: Convergence of FP when starting with (a, x)

Week 3 Homework

You can find more detailed descriptions of homework tasks in the GitHub repository.

1. Strategy profile evaluation ($\delta_i(\pi)$, $NashConv(\pi)$, $Exploitability(\pi)$)
2. Fictitious play
3. Exploitability convergence plots