Learning in Repeated Games

Prof. Jun Wang

Computer Science, UCL

Recap

- Lecture 1: Multiagent AI and basic game theory
- Lecture 2: Potential games, and extensive form and repeated games
- Lecture 3: Solving (“Learning”) Nash Equilibria
- Lecture 4: Bayesian Games, auction theory and mechanism design
- Lecture 5: Learning and deep neural networks
- Lecture 6: Single-agent Learning (1)
- Lecture 7: Multi-agent Learning (1)
- Lecture 8: Single-agent Learning (2)
- Lecture 9: Multi-agent Learning (2)
- Lecture 10: Multi-agent Learning (3)

Recap

Given a two-player, two-action matrix game, we have the payoff matrices as follows:

$$\mathbf{R}^1 = \begin{bmatrix} r_{11}^1 & r_{12}^1 \\ r_{21}^1 & r_{22}^1 \end{bmatrix} \quad \mathbf{R}^2 = \begin{bmatrix} r_{11}^2 & r_{12}^2 \\ r_{21}^2 & r_{22}^2 \end{bmatrix}$$

then we have:

$$\begin{aligned} u^1 &= r_{11}^1 - r_{12}^1 - r_{21}^1 + r_{22}^1, \quad b^1 = r_{12}^1 - r_{22}^1 \\ u^2 &= r_{11}^2 - r_{12}^2 - r_{21}^2 + r_{22}^2, \quad b^1 = r_{21}^2 - r_{22}^2. \end{aligned}$$

We use α, β represents the strategy pair for two players.

Recap

Dynamics of Strategy Pair - IGA

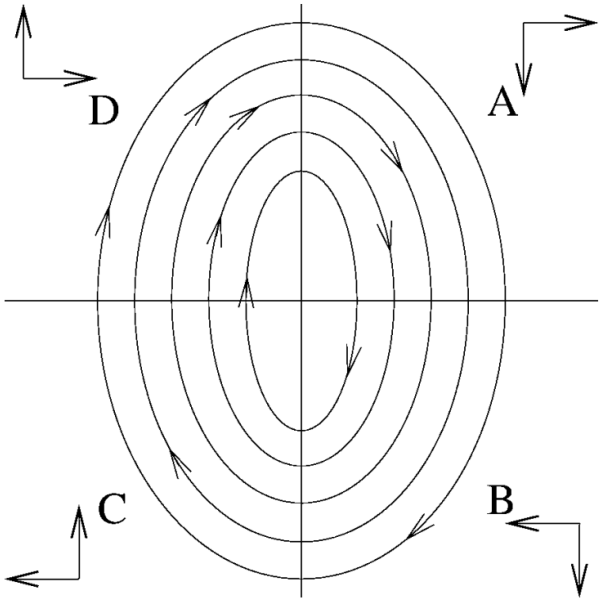
$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & u^1 \\ u^2 & 0 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} b^1 \\ b^2 \end{bmatrix}$$

Dynamics of Strategy Pair - WoLF IGA

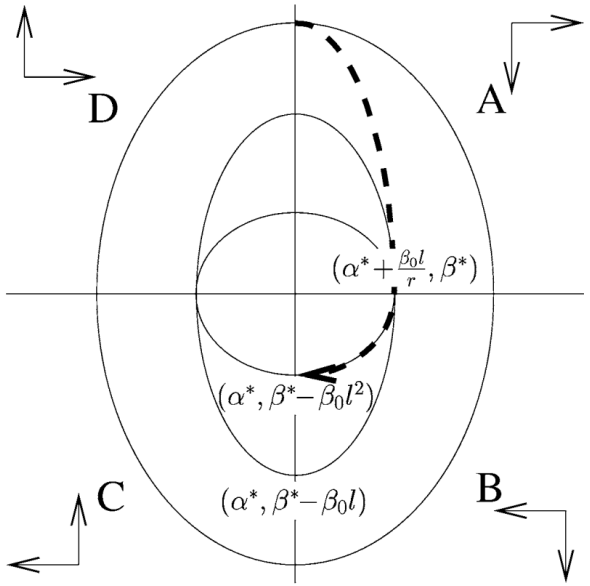
$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \eta^1(t)u^1 \\ \eta^2(t)u^2 & 0 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \eta^1(t)b^1 \\ \eta^2(t)b^2 \end{bmatrix}$$

Phase Portraits of IGA and WoLF IGA

IGA



WoLF IGA



The phase portraits of the IGA and the WoLF-IGA dynamics: when U has imaginary eigenvalues with negative real part.

Content

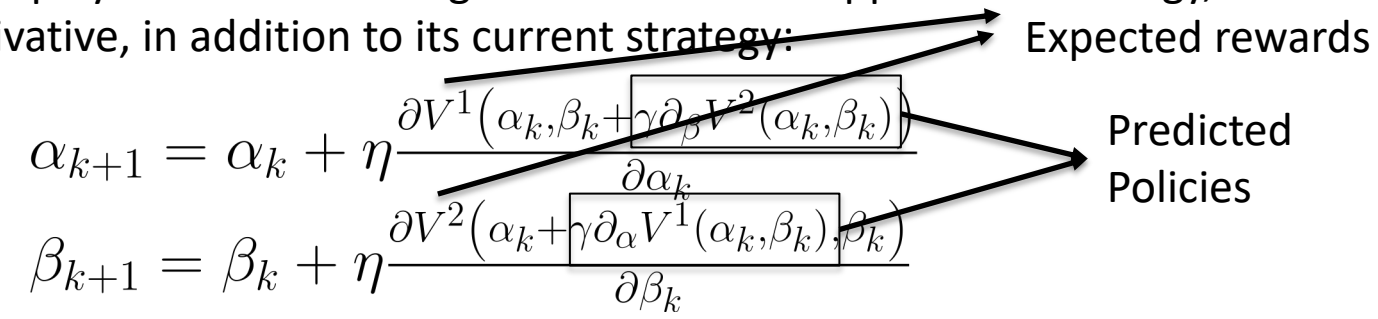
- IGA with Policy Prediction (based on Wen Yin's slides)
- Gradient Ascent Optimization as Dynamical Systems (based on Wen Yin's slides)
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

IGA with Policy Prediction (IGA-PP)

Suppose that one player knows its change direction of the opponent's strategy, i.e., strategy derivative, in addition to its current strategy:



$$\alpha_{k+1} = \alpha_k + \eta \frac{\partial V^1\left(\alpha_k, \beta_k + \gamma \partial_\beta V^2(\alpha_k, \beta_k)\right)}{\partial \alpha_k}$$

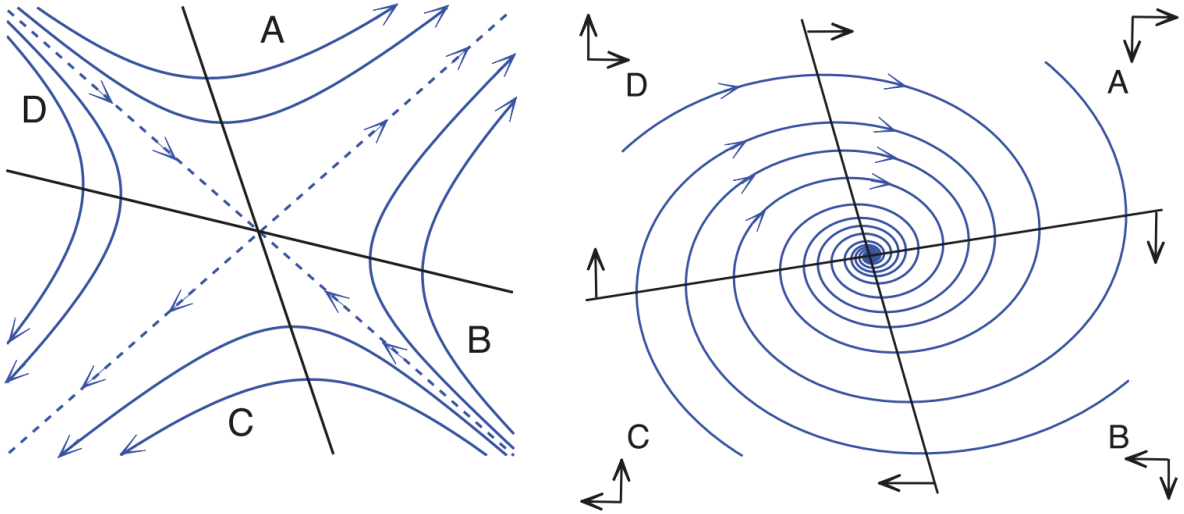
$$\beta_{k+1} = \beta_k + \eta \frac{\partial V^2\left(\alpha_k + \gamma \partial_\alpha V^1(\alpha_k, \beta_k), \beta_k\right)}{\partial \beta_k}$$

Dynamics of the strategy pair for IGA-PP is defined as following:

$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} \gamma u^1 u^2 & u^1 \\ u^2 & \gamma u^1 u^2 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \gamma u^1 b^2 + b^1 \\ \gamma u^2 b^1 + b^2 \end{bmatrix}$$

- Zhang, Chongjie, and Victor Lesser. "Multi-agent learning with policy prediction." Twenty-Fourth AAAI Conference on Artificial Intelligence. 2010.

Phase Portraits of IGA-PP



a) A saddle at the center

b) A stable focus at the center

The phase portraits of the IGA-PP dynamics: a) when U has real eigenvalues and b) when U has imaginary eigenvalues with negative real part.

2x2 Matrix Games as Dynamical Systems

A second order linear homogeneous system with constant coefficients can be given to describe the **2x2 Matrix Games Dynamics** :

$$\mathbf{X}' = U\mathbf{X}, \quad \text{where} \quad \mathbf{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad U = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}.$$

Recap we have dynamics for Gradient Ascend based methods

$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & u^1 \\ u^2 & 0 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} b^1 \\ b^2 \end{bmatrix}$$

IGA

$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \eta^1(t)u^1 \\ \eta^2(t)u^2 & 0 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \eta^1(t)b^1 \\ \eta^2(t)b^2 \end{bmatrix}$$

WoLF IGA

$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} \gamma u^1 u^2 & u^1 \\ u^2 & \gamma u^1 u^2 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \gamma u^1 b^2 + b^1 \\ \gamma u^2 b^1 + b^2 \end{bmatrix}$$

IGA-PP

2x2 Matrix Games as Dynamical Systems

A second order linear homogeneous system with constant coefficients can be given to describe the **2x2 Matrix Games Dynamics** :

$$\mathbf{X}' = U\mathbf{X}, \quad \text{where} \quad \mathbf{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad U = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}.$$

The equilibrium positions can be found by solving the stationary equation

$$U\mathbf{X} = \mathbf{0}.$$

This equation has the unique solution $\mathrm{X} = 0$ if the matrix U is **non-singular**, i.e. provided that det$U \neq 0$. In the case of a **singular matrix**, the system has an infinite number of equilibrium points.

Different Types of Equilibrium Points

Matrix U is non-singular:

Classification of equilibrium points is determined by **eigenvalues** λ_1,λ_2 of the matrix U.

The numbers λ_1,λ_2 can be found by solving the **auxiliary equation**:

#	Equilibrium Point	Eigenvalues λ_1,λ_2
1	Node	λ_1,λ_2 are real numbers of the same sign $(\lambda_1 \cdot \lambda_2 > 0)$
2	Saddle	λ_1,λ_2 are real numbers and non-zero of opposite sign $(\lambda_1 \cdot \lambda_2 < 0)$
3	Focus	λ_1,λ_2 are complex numbers, the real parts are equal and non-zero $(\mathrm{Re}\,\lambda_1 = \mathrm{Re}\,\lambda_2 \neq 0)$
4	Center	λ_1,λ_2 are purely imaginary numbers $(\mathrm{Re}\,\lambda_1 = \mathrm{Re}\,\lambda_2 = 0)$

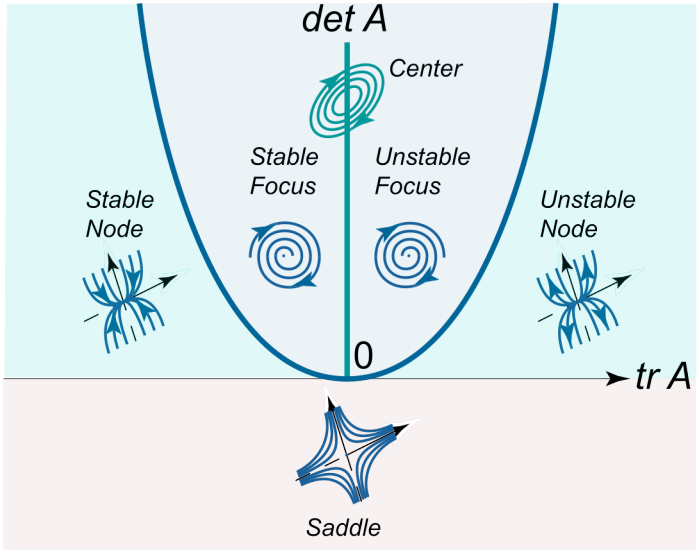
$$\lambda^2 - (u_{11} + u_{22})\,\lambda + u_{11}a_{22} - u_{12}u_{21} = 0.$$

In the case of purely imaginary roots (when the equilibrium point is a **centre**), we are dealing with the classical **stability in the sense of Lyapunov**.

https://www.math24.net/linear-autonomous-systems-equilibrium-points/

Phase Portraits of Equilibrium Points

#	Equilibrium Point	Eigenvalues λ_1, λ_2
1	Node	λ_1, λ_2 are real numbers of the same sign $(\lambda_1 \cdot \lambda_2 > 0)$
2	Saddle	λ_1, λ_2 are real numbers and non-zero of opposite sign $(\lambda_1 \cdot \lambda_2 < 0)$
3	Focus	λ_1, λ_2 are complex numbers, the real parts are equal and non-zero $(\mathrm{Re}\ \lambda_1 = \mathrm{Re}\ \lambda_2 \neq 0)$
4	Center	λ_1, λ_2 are purely imaginary numbers $(\mathrm{Re}\ \lambda_1 = \mathrm{Re}\ \lambda_2 = 0)$



https://www.math24.net/linear-autonomous-systems-equilibrium-points/

Lyapunov Functions

Let a function $V(\boldsymbol{X})$ be continuously differentiable in a neighbourhood δ of the origin. The function $V(\boldsymbol{X})$ is called the **Lyapunov function** for an autonomous system: $\boldsymbol{X}' = f(\boldsymbol{X})$

if the following conditions are met:

1. $V(\boldsymbol{X}) > 0 \; for \; all \; \boldsymbol{X} \in \delta \backslash \{0\}$;
2. $V(0) = 0$;
3. $\frac{dV}{dt} \leq 0 \; for \; all \; \boldsymbol{X} \in \delta$.

Lyapunov Stability Theorems

Theorem on stability in the sense of Lyapunov.
If in a neighbourhood δ of the zero solution $\boldsymbol{X} = 0$ of an autonomous system there is a Lyapunov function $V(\boldsymbol{X})$, then the equilibrium point $\boldsymbol{X} = 0$ of the system is Lyapunov stable.

Theorem on asymptotic stability.
If in a neighbourhood δ of the zero solution $\boldsymbol{X} = 0$ of an autonomous system there is a Lyapunov function, $\boldsymbol{X} = 0$ with a negative definite derivative $\frac{dV}{dt} < 0$ for all $\boldsymbol{X} \in \delta \backslash \{0\}$, then the equilibrium point $\boldsymbol{X} = 0$ of the system is asymptotically stable.

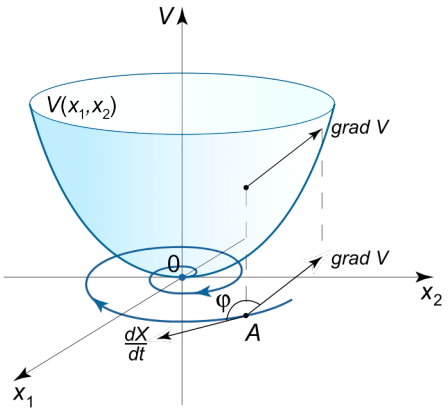
Note, the total derivative $\frac{dV}{dt}$ must be strictly negative (negative definite) in a neighbourhood of the origin for the asymptotic stability of the zero solution.

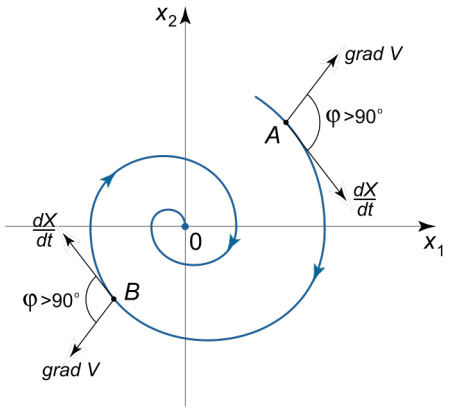
Lyapunov Stability Theorems

$$\frac{dV}{dt} = \frac{\partial V}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial V}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial V}{\partial x_n}\frac{dx_n}{dt}$$

Consider the case when the derivative of $V(\boldsymbol{X})$ in a neighbourhood δ of the origin is negative: $\frac{dV}{dt} = \left(grad\ V, \frac{d\boldsymbol{X}}{dt}\right) < \mathbf{0}$.

This means that the angle φ between the gradient vector and the velocity vector is greater than 90°. For a function of two variables, it is shown schematically:





https://www.math24.net/method-lyapunov-functions/

Case Study: Lyapunov Stability Analysis

Given two players matching pennies game:

$$\mathbf{R}^1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \mathbf{R}^2 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

We use the method of Lyapunov functions for the stability analysis to examine the convergence of Gradient Ascent Methods.

Let the Lyapunov function $V(x, y)$ have the form:

$$V(x, y) = x^2 + y^2$$

Lyapunov Stability - IGA

$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & u^1 \\ u^2 & 0 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} b^1 \\ b^2 \end{bmatrix} \quad \Rightarrow \quad \begin{aligned} \frac{\mathrm{d}\alpha}{\mathrm{d}t} &= u^1\beta + b^1, \\ \frac{\mathrm{d}\beta}{\mathrm{d}t} &= u^2\alpha + b^2. \end{aligned}$$

We set $\alpha = x + 0.5$ and $y = \beta + 0.5$, we can have

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= u^1(y + 0.5) + b^1 = 4y, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= u^2(x + 0.5) + b^2 = -4x. \end{aligned}$$

Lyapunov Stability - IGA

We use the method of Lyapunov functions for the stability analysis. Let the function $V(x, y)$ have the form:

$$V(x, y) = x^2 + y^2$$

We set $\alpha = x + 0.5$ and $y = \beta + 0.5$, we can have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = u^1(y + 0.5) + b^1 = 4y,$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = u^2(x + 0.5) + b^2 = -4x.$$

Lyapunov Stability - IGA

$$
\begin{aligned}
\frac{\mathrm{d}V}{\mathrm{d}t} &= \frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{\mathrm{y}}{\mathrm{t}} \\
&= 2x \cdot (u^1(y + 0.5) + b^1) + 2y \cdot (u^2(x + 0.5) + b^2) \\
&= 2x \cdot 4y + 2y \cdot (-4x) \\
&= 8xy - 8xy \\
&\equiv 0
\end{aligned}
$$

not stable

Lyapunov Stability – IGA-PP

$$\begin{bmatrix} \partial\alpha/\partial t \\ \partial\beta/\partial t \end{bmatrix} = \underbrace{\begin{bmatrix} \gamma u^1 u^2 & u^1 \\ u^2 & \gamma u^1 u^2 \end{bmatrix}}_{U} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \gamma u^1 b^2 + b^1 \\ \gamma u^2 b^1 + b^2 \end{bmatrix} \quad \Rightarrow \quad \begin{aligned} \frac{\mathrm{d}\alpha}{\mathrm{d}t} &= \gamma u^1 u^2 \alpha + u^1 \beta + \gamma u^1 b^2 + b^1, \\ \frac{\mathrm{d}\beta}{\mathrm{d}t} &= \gamma u^1 u^2 \beta + u^2 \alpha + \gamma u^2 b^1 + b^2. \end{aligned}$$

We set $\alpha = x + 0.5$ and $y = \beta + 0.5$, we can have

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \gamma u^1 u^2 x + u^1 (y + 0.5) + \gamma u^1 b^2 + b^1 = -16\gamma x + 4y, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \gamma u^1 u^2 y + u^2 (x + 0.5) + \gamma u^2 b^1 + b^2 = -16\gamma y - 4x. \end{aligned}$$

Lyapunov Stability – IGA-PP

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}t} &= \frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} \\ &= 2x \cdot (-16\gamma x + 4y) + 2y \cdot (-16\gamma y - 4x) \\ &= -32\gamma(x^2 + y^2) < 0\end{aligned}$$

asymptotic stability

Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Learning in repeated game

- Players/agents in classical game theory (previous lectures) have
  - a perfect knowledge of the environment and
  - the payoff tables, and
  - try to maximize their individual payoff.
- Thus, the goal is to figure out, a priori, how to optimize its actions, e.g., calculate Nash equilibria
- However, when information is incomplete or in a repeated game, it becomes impossible to judge what choices are the most rational
- The question then facing a player/agent becomes how to learn to optimize its behaviour and maximize its return, based on local knowledge and through a process of trial and error.

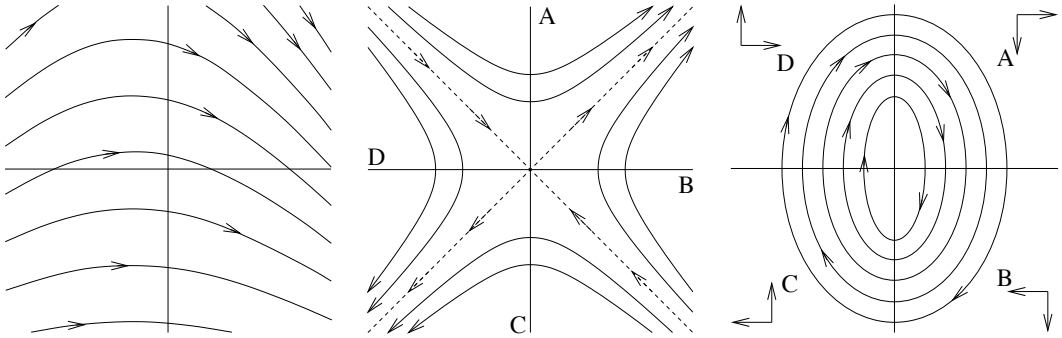
Tuyls, Karl, and Simon Parsons. "What evolutionary game theory tells us about multiagent learning." *Artificial Intelligence*171.7 (2007): 406-416.

learning in single agent

- A typical AI concerns the learning performed by an individual agent
- In that setting, the goal is to design an agent that learns to function successfully in an environment that is unknown and potentially also changes as the agent is learning
  - Learning to recommend in collaborative filtering
  - Learning to predict click-through rate by logistic regression

learning over multiple agents

- In a multi-agent (player) setting, the environment contains other agents (players) – (we are going to use term “player” and “agent” interchangeably)
- Additional complication:
  - the learning of other agents will change the environment, thus making an impact on the learning of our player, and
  - The learning of our agent will also influence the learning of other agents
- The simultaneous learning of them means that
  - every learning rule leads to a dynamical system, and
  - sometimes even very simple learning rules can lead to complex global behaviours of the system



Bowling, Michael, and Manuela Veloso. "Multiagent learning using a variable learning rate." Artificial Intelligence 136.2 (2002): 215-250.

Interaction between learning and teaching

- Also multi-agent systems cannot separate the phenomenon of *learning* from that of *teaching*
- When choosing a course of action, a player must take into account
  - not only what he has learned from other player' past behaviour,
  - but also how he wishes to influence their future behaviour

Image source: https://rryshke.files.wordpress.com/2012/11/art-of-teaching.jpg

An infinitely repeated game

- A repeated game: a given game (e.g., in normal form) is played multiple times by the same set of players.
  - The game being repeated is called the *stage game*.
- Infinitely Repeated Game: the stage game is infinitely played
- In IRG, average reward is
  - the payoff to a given player is the *limit average* of his payoffs in the individual stage games

	L	R
T	$1, 0$	$3, 2$
B	$2, 1$	$4, 0$

Stackelberg game as the stage game

Given an infinite sequence of payoffs $r^{(1)}, r^{(2)}, \ldots$ for player i, the average reward of i is

$$\lim_{k \to \infty} \frac{\sum_{j=1}^{k} r_i^{(j)}}{k}.$$

An infinitely repeated game

- (B,L) is the unique Nash equilibrium of the game
  - Agent 1 (the row player) has a dominant strategy, B
- Observations:
  - If agent 1 were to play B repeatedly, it is reasonable to expect that agent 2 would always respond with L.

	L	R
T	$1, 0$	$3, 2$
B	$2, 1$	$4, 0$

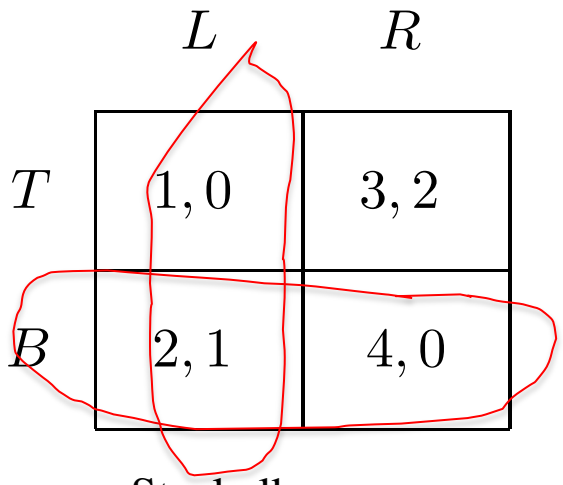
Stackelberg game as the stage game

Given an infinite sequence of payoffs $r^{(1)}, r^{(2)}, \ldots$ for player i, the average reward of i is

$$\lim_{k \to \infty} \frac{\sum_{j=1}^{k} r_i^{(j)}}{k}.$$

An infinitely repeated game

- (B,L) is the unique Nash equilibrium of the game
  - Agent 1 (the row player) has a dominant strategy, B
- Observations:
  - If agent 1 were to play B repeatedly, it is reasonable to expect that agent 2 would always respond with L.



Stackelberg game as the stage game

Given an infinite sequence of payoffs $r^{(1)}, r^{(2)}, \ldots$ for player i, the average reward of i is

$$\lim_{k \to \infty} \frac{\sum_{j=1}^{k} r_i^{(j)}}{k}.$$

An infinitely repeated game

- (B,L) is the unique Nash equilibrium of the game
  - Agent 1 (the row player) has a dominant strategy, B
- Observations:
  - if agent 1 were to choose T instead, then agent 2's best response would be R, yielding a payoff large than that in Nash equilibrium

	L	R
T	$1, 0$	$3, 2$
B	$2, 1$	$4, 0$

Stackelberg game as the stage game

Given an infinite sequence of payoffs $r^{(1)}, r^{(2)}, \ldots$ for player i, the average reward of i is

$$\lim_{k \to \infty} \frac{\sum_{j=1}^{k} r_i^{(j)}}{k}.$$

An infinitely repeated game

- (B,L) is the unique Nash equilibrium of the game
  - Agent 1 (the row player) has a dominant strategy, B
- Observations:
  - if agent 1 were to choose T instead, then agent 2's best response would be R, yielding a payoff large than that in Nash equilibrium

	L	R
T	$1, 0$	$3, 2$
B	$2, 1$	$4, 0$

Stackelberg game as the stage game

Given an infinite sequence of payoffs $r^{(1)}, r^{(2)}, \ldots$ for player i, the average reward of i is

$$\lim_{k \to \infty} \frac{\sum_{j=1}^{k} r_i^{(j)}}{k}.$$

Teaching

- In a single-stage game it would be hard for agent 1 to convince agent 2 that he will play T , since B is a strictly dominated strategy
- However, in a repeated-game setting, agent 1 has an opportunity being a teacher
  - agent 1 could repeatedly play T; presumably, after a while agent 2, if he has any sense at all, would get the message and start responding with R

	L	R
T	$1, 0$	$3, 2$
B	$2, 1$	$4, 0$

Stackelberg game as the stage game

Given an infinite sequence of payoffs $r^{(1)}, r^{(2)}, \ldots$ for player i, the average reward of i is

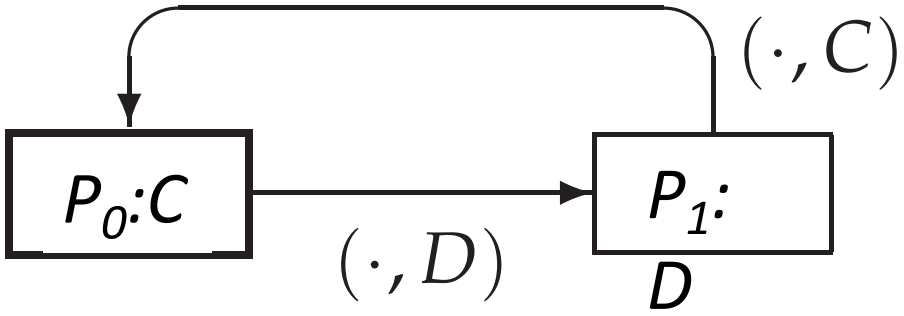
$$\lim_{k \to \infty} \frac{\sum_{j=1}^{k} r_i^{(j)}}{k}.$$

What constitutes learning?

- A repeated game is regarded as a nature setting for "learning"
  - temporal nature and
  - the regularity across time (at each time the same players are involved, and they play the same game as before)
- This allows us to consider strategies:

future action is selected based on the experience gained so far

  - The Tit-for-Tat (TfT) and trigger strategies (studied in repeated Prisoner's Dilemma) can be viewed as a rudimentary form of learning strategies



What constitutes learning?

- More complex strategies: an agent's next choice depends on the history of play in more sophisticated ways, e.g.,
  - the agent could guess that
  - the frequency of actions played by his opponent in the past might be his current mixed strategy, and
  - play a best response to that mixed strategy
- This basic learning rule is called *fictitious play*

What games require learning

- Repeated game
- Population game (will be explained shortly)
- Stochastic game (as we explained previously)

What are settings for learning

- Whether the game is *known* by the players
  - If the game is known, any "learning" that takes place is only about the strategies employed by the others
  - If the game is unknown, the agent can in addition learn about the structure of the game itself
- For instance, the agent may start out not knowing the payoff functions at a given stage game or additionally the transition probabilities (in a stochastic game setting), but learn those over time in the course of playing the game.
  - With certain learning strategies, agents can sometimes converge to an equilibrium even without knowing the game being played!
- Whether the game is *observable* by the players
  - do the players see each others' actions, and/or each others' payoffs?

Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Fictitious Play

- *Fictitious play* is a simple sequential procedure that learn the value of a game
- It is an instance of model-based learning,
  - the learner explicitly maintains beliefs about the opponent's strategy. The learning structure:

Initialize beliefs about the opponent's strategy
repeat
  Play a best response to the assessed strategy of the opponent
  Observe the opponent's actual play and update beliefs accordingly

Note that in this setting,

- the agent does not know the payoffs and payoff functions by other agents, and
- he, however, knows his own payoff matrix in the stage game (i.e., the payoff he would get in each action profile, whether or not encountered in the past).

Brown, George W. "Iterative solution of games by fictitious play." *Activity analysis of production and allocation* 13.1 (1951): 374-376.

Fictitious Play

- In fictitious play, an agent believes that
  - his opponent is playing the mixed strategy that is consistent with the empirical distribution of the opponent's previous actions
- Formally,
  - A is the set of the opponent's actions, and
  - for every a ∈ A, let w(a) be the number of times that the opponent has played action a.
  - Then, the agent assesses the opponent's mixed strategy as

$$P(a) = \frac{w(a)}{\sum_{a' \in A} w(a')}.$$

Fictitious Play

- Fictitious play is sensitive to the players' initial beliefs or prior
  - which can be interpreted as action counts that were observed before the start of the game
  - Note that one must pick some nonempty prior belief for each agent; the prior beliefs cannot be (0, . . . , 0) since this does not define a meaningful mixed strategy
- The prior beliefs can have a radical impact on the learning process
- Drawback: in fictitious play each agent assumes a stationary policy of the opponent,
  - yet no agent plays a stationary policy except when the process happens to converge to one!

Fictitious Play: an example

- In a repeated Prisoner's Dilemma game,
  - if the opponent has played
  - C, C, D, C, D in the first five games,

	C	D
C	$2,2$	$0,3$
D	$3,0$	$1,1$

- We can represent a player's beliefs with either a probability measure or with the set of counts $(w(a_1), \ldots, w(a_k))$
  - before the sixth game he is assumed to be playing the mixed strategy (w(C)=0.6, w(D)=0.4)
  - In the sixth game, what would be the best response to (w(C)=0.6, w(D)=0.4)?

Fictitious Play: an example

- In a repeated Prisoner's Dilemma game,
  - if the opponent has played
  - C, C, D, C, D in the first five games,

	C	D
C	$2,2$	$0,3$
D	$3,0$	$1,1$

- we can represent a player's beliefs with either a probability measure or with the set of counts $(w(a_1), \ldots, w(a_k))$
  - before the sixth game he is assumed to be playing the mixed strategy (w(C)=0.6, w(D)=0.4)
  - In the sixth game, what would be the best response to (w(C)=0.6, w(D)=0.4)?

If C were chosen, the reward is 2x0.6+0x0.4=1.2
If D were chosen, the reward is 3x0.6+1x0.4=2.2

Fictitious Play: an example

- Two players are playing a repeated game of Matching Pennies.
- Each player is using fictitious play learning to update his beliefs and select actions.
  - Player 1 begins the game with the prior belief that player 2 has played heads 1.5 times and tails 2 times
  - Player 2 begins with the prior belief that player 1 has played heads 2 times and tails 1.5 times
- How will the players play?

	Heads	Tails
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Matching Pennies game

Fictitious Play: an example

- Each player ends up alternating back and forth between playing heads and tails.
- As the number of rounds tends to infinity, the empirical distribution of the play of each player will converge to (0.5, 0.5).
- If we take this distribution to be the mixed strategy of each player, the play converges to the unique Nash equilibrium of the normal form stage game : each player plays the mixed strategy (0.5, 0.5)

	Heads	Tails
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Matching Pennies game

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Fictitious Play: an example

- Each player ends up alternating back and forth between playing heads and tails.
- As the number of rounds tends to infinity, the empirical distribution of the play of each player will converge to (0.5, 0.5).
- If we take this distribution to be the mixed strategy of each player, the play converges to the unique Nash equilibrium of the normal form stage game : each player plays the mixed strategy (0.5, 0.5)

	Heads	Tails
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Matching Pennies game

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Fictitious Play: an example

- Each player ends up alternating back and forth between playing heads and tails.
- As the number of rounds tends to infinity, the empirical distribution of the play of each player will converge to (0.5, 0.5).
- If we take this distribution to be the mixed strategy of each player, the play converges to the unique Nash equilibrium of the normal form stage game : each player plays the mixed strategy (0.5, 0.5)

Heads: 1x1.5+(-1)x2

	Heads	Tails
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Tails: (-1)x1.5+1x2

Matching Pennies game

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Fictitious Play: an example

- Each player ends up alternating back and forth between playing heads and tails.
- As the number of rounds tends to infinity, the empirical distribution of the play of each player will converge to (0.5, 0.5).
- If we take this distribution to be the mixed strategy of each player, the play converges to the unique Nash equilibrium of the normal form stage game : each player plays the mixed strategy (0.5, 0.5)

Heads: 1x1.5+(-1)x2

	Heads	Tails
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Matching Pennies game

Tails: (-1)x1.5+1x2

Tails>Heads

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Some properties

- Steady state: an action profile a is a steady state of fictitious play
  - if it is the case that whenever a is played at round t it is also played at round t + 1 (and hence in all future rounds as well)
- A tight connection between steady states and pure-strategy Nash equilibria:
  - **Theorem 1** *If a pure-strategy profile is a strict Nash equilibrium of a stage game, then it is a steady state of fictitious play in the repeated game*
  - Note that the pure-strategy profile must be a *strict* Nash equilibrium, i.e.,
    - no agent can deviate to another action without strictly decreasing its payoff

Some properties

- Steady state: an action profile a is a steady state of fictitious play
  - if it is the case that whenever a is played at round t it is also played at round t + 1 (and hence in all future rounds as well)
- A tight connection between steady states and pure-strategy Nash equilibria:
  - **Theorem 2** *If a pure-strategy profile is a steady state of fictitious play in the repeated game, then it is a (possibly weak) Nash equilibrium in the stage game.*
  - Note that fictitious play may not always converges to a Nash equilibrium,
    - as agents can only play pure strategies and a pure-strategy Nash equilibrium may not exist in a given game

Some properties

- However, while the stage game strategies may not converge, the empirical distribution of the stage game strategies may
- This was the case in the Matching Pennies example,
  - where the empirical distribution of the each player's strategy converged to their mixed strategy in the (unique) Nash equilibrium of the game.
- The following theorem shows that this was no accident.
  - **Theorem 3** *If the empirical distribution of each player's strategies converges in fictitious play, then it converges to a Nash equilibrium.*

Fictitious Play: an example

- However, although the theorem gives sufficient conditions for the empirical distribution to converge to a mixed equilibrium, no claims made about the distribution of the particular actions played
- To see this, consider a *repeated Anti-Coordination game* here
  - two pure Nash equilibria of this game, (A, B) and (B, A), and one mixed Nash equilibrium: each agent mixes A and B with probability 0.5
  - Either of the two pure-strategy equilibria earns each player a payoff of 1, and the mixed-strategy equilibrium earns each player a payoff of 0.5

	A	B
A	$0, 0$	$1, 1$
B	$1, 1$	$0, 0$

The Anti-Coordination game as the stage game.

How the fictitious play is conducted if we assume that the weight function for each player is initialized to (1, 0.5)?

Fictitious Play: an example

- In fictitious play, we assume that the weight function for each player is initialized to (1, 0.5)
- The play of each player converges to the mixed strategy Nash equilibrium (0.5, 0.5)
- However, the payoff received by each player is 0,
  - since the players never hit the outcomes with positive payoff.
- It shows that although the empirical distribution of the strategies converges to the mixed strategy Nash equilibrium,
- the players may not receive the expected payoff of the Nash equilibrium,
  - because their actions are miscorrelated!

	A	B
A	$0, 0$	$1, 1$
B	$1, 1$	$0, 0$

The Anti-Coordination game as the stage game.

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1,0.5)	(1,0.5)
1	B	B	(1,1.5)	(1,1.5)
2	A	A	(2,1.5)	(2,1.5)
3	B	B	(2,2.5)	(2,2.5)
4	A	A	(3,2.5)	(3,2.5)
⋮	⋮	⋮	⋮	⋮

Fictitious play of a repeated Anti-Coordination game.

Fictitious Play: an example

- *The empirical distributions of players' actions need not converge at all*.
  - Consider the game, due to Shapley, a modification of the rock-paper-scissors game; this game is not zero sum.
  - The unique Nash equilibrium of this game is for each player to play the mixed strategy (1/3, 1/3, 1/3)
- In fictitious play, player 1's weight function is initialized to (0, 0, 0.5) and player 2's weight function is initialized to (0,0.5,0).

	Rock	Paper	Scissors
Rock	0, 0	0, 1	1, 0
Paper	1, 0	0, 0	0, 1
Scissors	0, 1	1, 0	0, 0

Shapley's Almost-Rock-Paper-Scissors game as the stage game

Fictitious Play: an example

- *The empirical distributions of players' actions need not converge at all.*
  - Consider the game, due to Shapley, a modification of the rock-paper-scissors game; this game is not zero sum.
  - The unique Nash equilibrium of this game is for each player to play the mixed strategy (1/3, 1/3, 1/3)
- In fictitious play, player 1's weight function is initialized to (0, 0, 0.5) and player 2's weight function is initialized to (0,0.5,0).
- The play of this game is shown on the right.
- Although it is not obvious from these first few rounds, it can be shown that the empirical play of this game never converges to any fixed distribution.

	Rock	Paper	Scissors
Rock	0, 0	0, 1	1, 0
Paper	1, 0	0, 0	0, 1
Scissors	0, 1	1, 0	0, 0

Shapley's Almost-Rock-Paper-Scissors game as the stage game

Round	1's action	2's action	1's beliefs	2's beliefs
0			(0,0,0.5)	(0,0.5,0)
1	Rock	Scissors	(0,0,1.5)	(1,0.5,0)
2	Rock	Paper	(0,1,1.5)	(2,0.5,0)
3	Rock	Paper	(0,2,1.5)	(3,0.5,0)
4	Scissors	Paper	(0,3,1.5)	(3,0.5,1)
5	Scissors	Paper	(0,1.5,0)	(1,0,0.5)
⋮	⋮	⋮	⋮	⋮

Fictitious play of a repeated game of the Almost-Rock-Paper-Scissors game.

Fictitious Play: conclusions

- It is interesting not because it is realistic or has strong guarantees, but because
  - It is very simple to state and
  - gives rise to nontrivial properties
- But it is very limited;
  - its model of beliefs and belief update is mathematically constraining, and
  - is clearly implausible as a model of human learning
- There exist various variants of fictitious play that are somewhat better, such as *smoothed fictitious play*

One of the many applications

Iterative Computation of Cournot Equilibrium*

LARS THORLUND-PETERSEN

Norwegian School of Economics and Business Administration,
N-5035 Bergen-Sandviken, Norway

Received October 18, 1988

In a homogeneous Cournot model with quasi-concave profit functions the problem of determining an equilibrium can be posed as one of solving an equation in one real variable: total sales. If the response functions are monotone or firms are identical, then a certain iterative process based on averaging converges to an equilibrium. Such iterations have the interpretation that every firm responds to the average of sales by other firms in previous periods. *Journal of Economic Literature* Classification Number: 026.

1. INTRODUCTION

Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Smoothed Fictitious Play

- Mathematically, Fictitious Play adopts at time *t+1* a pure strategy s_i that

$$\max_{s_i} u_i(s_i, P^t)$$

where P^t is the empirical distribution of opponent's play until time t. *u_i is the expected utility*

- Smoothed Fictitious Play, instead of playing the best response to the empirical frequency, introduces a perturbation that gradually diminishes over time
  - agent i adopts a mixed strategy σ_i that maximizes

$$\operatorname*{argmax}_{\sigma_i} \sum_{s_i} \sigma_i(s_i)\, u_i(s_i, P^t) - \beta v_i(\sigma_i)$$

Where β is any constant, and $v_i(\sigma_i)$ can be the entropy function $v_i(\sigma_i) = \sum_{s_i} \sigma_i(s_i) \log \sigma_i(s_i)$

Fudenberg, Drew, and David K. Levine. "Consistency and cautious fictitious play." *Journal of Economic Dynamics and Control* 19.5-7 (1995): 1065-1089.

Smoothed Fictitious Play

- The first order condition for the maximum gives

$$u_i(s_i,P^t) - \beta \log \sigma_i(s_i) + \lambda = 0$$

where λ is the Lagrange multiplier corresponding to the constraint that the probabilities $\sigma_i()$ a must sum to one

- Solving it gives:

$$\sigma_i(s_i) = \frac{\exp(u_i(s_i,Pt)/\beta)}{\sum_{s_i'} \exp(u_i(s_i',Pt)/\beta)}$$

- It allows a more satisfactory explanation for convergence to mixed-strategy equilibria in fictitious play-like models.
  - For example, in matching pennies the per-period play can actually converge to the mixed strategy equilibrium.
  - In addition, SFP avoids the discontinuity inherent in standard fictitious play, where a small change in the data can lead to an abrupt change in behaviour.
  - With SFP, if beliefs converge, play does too.

Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Rational learning

- *Rational learning*,aka *Bayesian learning,* adopts the same general model-based scheme as fictitious play
- Unlike fictitious play, however, it allows players to have a much richer set of beliefs about opponents' strategies:
  - In fictitious play, strategies are limited to ones derived from the stage game (only conditional on the empirical distribution of opponent's actions)
  - But in rational learning, the set of strategies comes from the entire repeated-game, conditional on the history plays, e.g., TfT in repeated Prisoner's Dilemma
- Thus, the beliefs of each player about his opponent's strategies may be expressed by any probability distribution over the set of all possible strategies

Kalai, Ehud, and Ehud Lehrer. "Rational learning leads to Nash equilibrium." *Econometrica: Journal of the Econometric Society* (1993): 1019-1045.

Rational learning

- Similar to fictitious play, each player begins the game with some prior beliefs.
- After each round, the player uses Bayesian inference to update their beliefs
- The Bayesian update for opponent's playing a particular strategy:

Likelihood from the history

Prior distribution

$$P_i(s_{-i}|h) = \frac{P_i(h|s_{-i})P_i(s_{-i})}{\sum_{s'_{-i} \in S^i_{-i}} P_i(h|s'_{-i})P_i(s'_{-i})}.$$

where

- S^i_{-i} denotes the set of the opponent's strategies considered possible by player i, and $s_{-i} \in S^i_{-i}$,
- H denotes the set of possible histories of the game, and
- h $\in H$

Rational learning

- Recall *grim trigger strategy in the* infinitely repeated Prisoner's Dilemma game :
  - choose *C* so long as the other player chooses *C*;
  - if in any period the other player chooses *D*, then choose *D* in *every subsequent period*
- A general case: *limited punishment* g^T:
  - choose *C* so long as the other player chooses *C*;
  - if in any period the other player chooses *D*, then choose *D* in in the following T times and goes back to *C*.
- A rational learning setting:
  - the strategy space consists of the strategies g^1, g^2, . . . g^T ,…,g^∞ ; g^∞ is the *trigger strategy*
  - each player happens to select a best response from among g^0,g^1,…,g^∞.
- After playing each round of the repeated game, each player performs Bayesian updating

	C	D
C	$2,2$	$0,3$
D	$3,0$	$1,1$

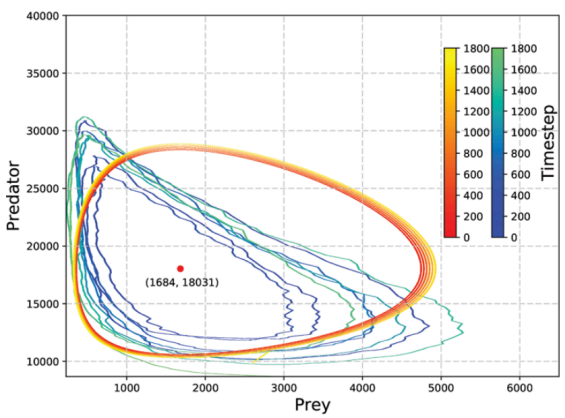
Prisoner's Dilemma as the stage game

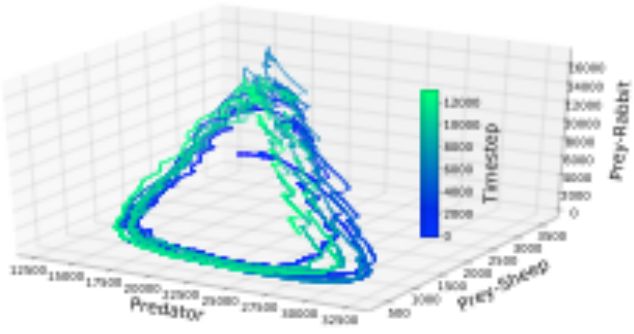
$$P_i(g_T|h_t) = \begin{cases} 0 & \text{if } T \leq t; \\ \frac{P_i(g_T)}{\sum_{k=t+1}^{\infty} P_i(g_k)} & \text{if } T > t. \end{cases}$$

if player i has observed that player j has always cooperated after history $h_t \in H$

Evolutionary learning in populations of agents

- Learning in a population of agents:
  - we mean the change in the constitution and behaviour of that population over time
- These models were originally developed by population biologists to model the process of biological evolution, and
- later adopted and adapted by other fields





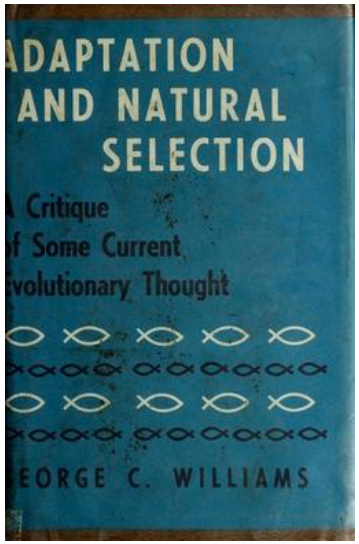
Yang, Yaodong, et al. "An Empirical Study of AI Population Dynamics with Million-agent Reinforcement Learning." *arXiv preprint arXiv:1709.04511* (2017).

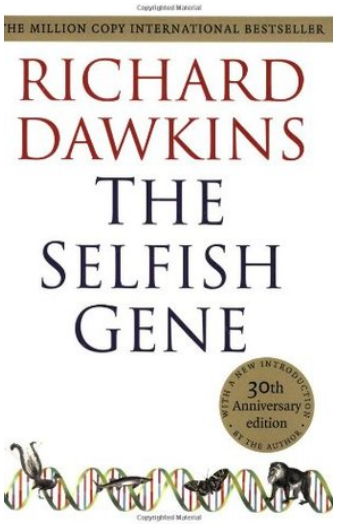
Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Background: evolutionary biology

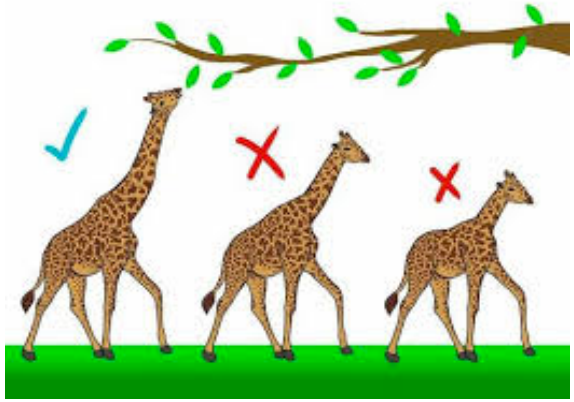
- Gene-centric view of evolution
  - An organism's genes largely determine its observable characteristics (fitness) in a given environment
    - More fit organisms will produce more offspring
  - This causes **genes** that provide greater fitness to increase their representation in the population via natural selection





1966 1976

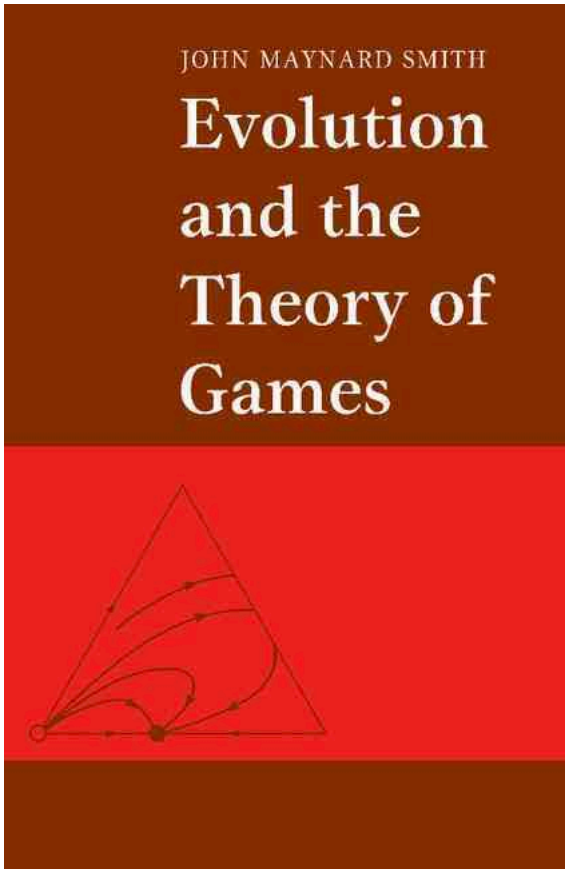
Gene-centric view of evolution



Natural selection

Evolutionary game theory

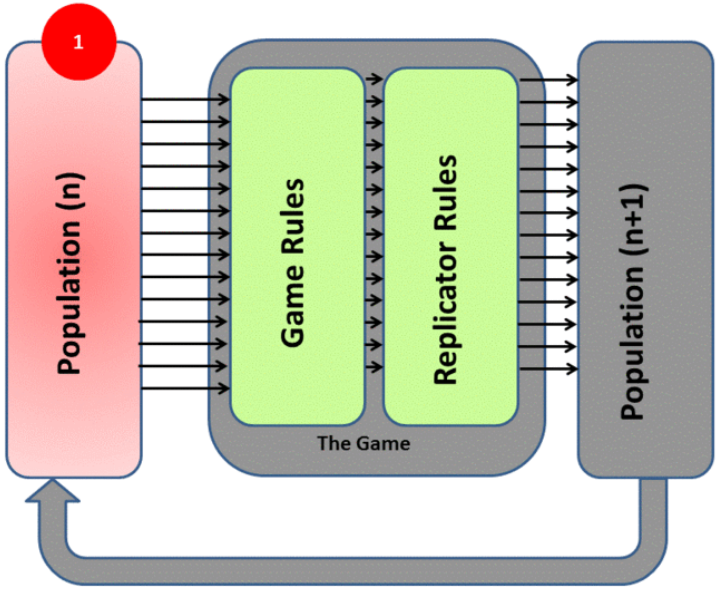
- In 1973 biologist John Maynard Smith and mathematician George R. Price showed how game theory applies to the behaviour of animals
- The idea of applying game theory to animals seemed strange at the time,
  - because game theory had always been about rationality
  - Animals hardly fit the bill
- Maynard Smith made three critical shifts from traditional game theory
  - strategy,
  - equilibrium, and
  - the nature of agent interactions



Maynard Smith's 1982 book has become a classic.

Background: Evolutionary Game theory

- Regular game theory
  - Individual players make decisions
  - Payoffs depend on decisions made by all
  - The reasoning about what other players might do happen simultaneously
- ***Evolutionary*** game theory
  - Game theory continues to apply even if no individual is reasoning or making explicit decisions
  - Decisions may thus not be conscious
  - What behavior will persist in a population?



Img source: https://en.wikipedia.org/wiki/Evolutionary_game_theory

Evolutionary game theory

- Key insight
  - Many behaviors involve the *interaction* of multiple organisms in a population
  - The success of an organism depends on how its behavior interacts with that of others
    - *Can‘t measure fitness of an individual organism along*
  - So fitness must be evaluated in the context of the full population in which it lives
- Analogous to game theory!
  - Organisms‘s genetically determined characteristics and behavior = Strategy
  - Fitness = Payoff
  - Payoff depends on strategies of organisms with which it interacts = Game matrix

Motivating example

- Let‘s look at a species of a beetle
  - Each beetle‘s fitness depends on finding and processing food effectively
  - Mutation introduced
    - Beetles with mutation have larger body size
    - Large beetles need more food
- What would we expect to happen?
  - Large beetles need more food
  - This makes them less fit for the environment
  - The mutation will thus die out over time
- But there is more to the story...

Motivating example

- Beetles compete with each other for food
  - Large beetles more effective at claiming *above-average* share of the food
- Assume food competition is among pairs
  - Small vs. Small : get equal shares of food
  - Large vs. small: Large beetle gets the majority of food from Small beetle
  - Large vs. Large: get equal shares of food, but Large beetles always experience less fitness benefit from given quantity of food
    - Need to maintain their expensive metabolism (the chemical processes in their body)

Motivating example

- The body-size game between two beetles

	Small	Large
Small	5, 5	1, 8
Large	8, 1	3, 3

- Something funny about this
  - No beetle is asking itself: *"Do I want to be small or large?"*
- Need to think about strategy changes that operate over longer time scales
  - Taking place as shifts in population under evolutionary forces!

Evolutionarily stable strategies

- Suppose each beatle is repeatedly paired off with other beetles at random
  - Population large enough so that there are no repeated interactions between two beetles
- A beetle‘s fitness = average fitness from food interactions = reproductive success
  - More food thus means more offspring to carry genes (strategy) to the next generation

Evolutionary stable strategies

- The concept of a Nash equilibrium doesn‘t work in this setting
  - Nobody is changing their personal strategy
- Instead, we want an ***evolutionary stable strategy***
  - A genetically determined strategy that tends to persist once it is prevalent in a population
  - *Def*:
    - **A strategy is *evolutionarily stable* if everyone uses it, and any small group of invaders with a different strategy will die off over multiple generations**
- Need to make this precise...

Motivating example

- Is **Small** an evolutionarily stable strategy?
- Let's use the definition
  - Suppose for some small number ε, a 1- ε fraction of population use **Small** and ε use **Large**
  - In other words, **Large** beetles invades a population of **Small** beetles

	Small	Large
Small	5, 5	1, 8
Large	8, 1	3, 3

What is the expected payoff to a **Small** beetle in a random interaction?

With prob. 1- ε, meet another **Small** beetle for a payoff of 5

With prob. ε, meet **Large** beetle for a payoff of 1

Expected payoff: $5(1-\varepsilon) + 1\,\varepsilon = 5-4\,\varepsilon$

Motivating example

- Is **Small** an evolutionarily stable strategy?
- Let's use the definition
  - Suppose for some small number ε, a 1- ε fraction of population use **Small** and ε use **Large**
  - In other words, **Large** beetles invades a population of **Small** beetles

	Small	Large
Small	5, 5	1, 8
Large	8, 1	3, 3

What is the expected payoff to a **Large** beetle in a random interaction?

With prob. 1- ε, meet a **Small** beetle for payoff of 8

With prob. ε, meet another **Large** beetle for a payoff of 3

Expected payoff: $8(1-\varepsilon) + 3\varepsilon = 8-5\varepsilon$

Motivating example

- Expected fitness of **Large beetles** is 8-5 ε
- Expected fitness of **Small beetles** is 5-4 ε
  - For small enough ε (and even big ε), the fitness of **Large beetles** exceeds the fitness for **Small**
  - Thus **Small** is NOT evolutionarily stable
- What about the Large strategy?
  - Assume ε fraction are **Small**, rest **Large**.
  - Expected payoff to **Large**: $3(1-\varepsilon) + 8\varepsilon = 3+5\varepsilon$
  - Expected payoff to **Small**: $1(1-\varepsilon) + 5\varepsilon = 1+4\varepsilon$
  - **Large** is evolutionarily stable

Motivating example

- Summary
  - A few large beetles introduced into a population consisting of small beetles
  - Large beetles will do really well:
    - They rarely meet each other
    - They get most of the food in most competitions
  - Population of small beetles cannot drive out the large ones
    - So **Small** is not evolutionarily stable

Motivating example

- Summary
  - Conversely, a few small beetles will do very badly
    - They will lose almost every competition for food
  - A population of large beetles *resists* the invasion of small beetles
  - **Large** is thus evolutionarily stable
- The structure is like prisoner‘s dilemma
  - Competition for food = arms race
  - Beetles can‘t change body sizes, but evolutionarily forces over multiple generations are achieving analogous effect

Evolutionary arms races

- Lots of examples
  - Height of trees follows prisoner‘s dilemma
    - Only applies to a particular height range
    - More sunlight offset by fitness downside of height
  - Roots of soybean plants to claim resources
    - Conserve vs. Explore
- Hard to truly determine payoffs in real-world settings

Evolutionary arms races

- One recent example with known payoffs
  - Virus populations can play an evolutionary version of prisoner‘s dilemma
  - Virus A
    - Infects bacteria
    - Manifactures products required for replication
  - Virus B
    - Mutated version of A
    - Can replicate inside bacteria, but less efficiently
    - Benefits from presence of A
  - Is B evolutionarily stable?

Virus game

- Look at interactions between two viruses

	A	B
A	1.00, 1.00	0.65, 1.99
B	1.99, 0.65	0.83, 0.83

  - Viruses in a pure A population do better than viruses in pure B population
  - But regardless of what other viruses do, higher payoff to be B
- Thus B is evolutionarily stable
  - Even though A would have been better
  - Similar to the exam-presentation game

What happens in general?

- Under what conditions is a strategy evolutionarily stable?
  - Need to figure out the right form of the payoff matrix

Organism 2

Organism 1	S	T
S	*a*, *a*	*b*, *c*
T	*c*, *b*	*d*, *d*

  - How do we write the condition of evolutionary stability in terms of these 4 variables, a,b,c,d?

What happens in general?

- Look at the definition again
  - Suppose again that for some small number ε:
    - A 1- ε fraction of the population uses S
    - An ε fraction of the population uses T
- What is the payoff for playing S in a random interaction in the population?
  - Meet another S with prob. 1- ε. Payoff = a
  - Meet T with prob. ε. Payoff = b
  - Expected payoff = $a(1-\varepsilon)+b\,\varepsilon$
- Analogous for playing T
  - Expected payoff = $c(1-\varepsilon)+d\,\varepsilon$

What happens in general?

- Therefore, S is evolutionarily stable if for all small values of ε :
  - $a(1-\varepsilon)+b\,\varepsilon > c(1-\varepsilon)+d\,\varepsilon$
  - When ε is *really* small (goes to 0), this is
    - $a > c$
  - When $a=c$, the left hand side is larger when
    - $b > d$
- In other words
  - In a two-player, two-strategy symmetric game, S is evolutionarily stable when either
    - a > c, or
    - a = c, and b > d

What happens in general?

- Intuition
  - In order for S to be evolutionarily stable, then:
    - Using S against S must be *at least* as good as using T against S
    - Otherwise, an invader using T would have higher fitness than the rest of the population
  - If S and T are equally good responses to S
    - S can only be evolutionarily stable if those who play S do better against T than what those who play T do with each another
    - Otherwise, T players would do as well against the S part of the population as the S players

Relationship with Nash equilibria

- Let‘s look at Nash in the symmetric game

	S	T
S	a, a	b, c
T	c, b	d, d

  - When is (S,S) a Nash equilibrium?
  - S is a best response to S: $a \geq c$
- Compare with evolutionarily stable strategies:
  - (i) $a > c$ or (ii) $a = c$ and $b > d$
- Very similar!

Interpretation of mixed strategies

- Can interpret this in two ways
  - each agents plays the same mixed strategy, or
  - A fraction of the population playing each of the underlying pure strategies in proportion to its contribution to the mixed strategy
- As the stage game is a one-shot, it is rarely plausible to hold that an individual will play a strictly mixed strategy
- Thus, in general, the heterogeneous population interpretation is superior

ESS: definition of the stage game

- Consider a two-player normal form symmetric game:
  - both players have the set of pure strategies $S = \{s_1, \dots, s_n\}$
  - $u_1(s_i, s_j)$: the payoffs of an agent playing s_i when playing with another agent using s_j.
  - Symmetric payoff: $u_1(s_i, s_j) = u_2(s_i, s_j) \equiv u_{i,j}$
  - Symmetric in strategy: agents cannot condition their play on whether they are player 1 or player 2
  - $\mathrm{U} = (u_{i,j})$: the matrix of the symmetric game

ESS: how the stage game is played?

- Stage game G is a symmetric game with matrix U
- Large population of agents is to play the game
  - In each period t=1,2,…, agents are randomly paired and they play the stage game G once
  - Each agent has certain type i $\in \{1, \dots, n\}$, i.e., the agent uses strategy s_i in the stage game.
- The state of the game (also called the strategy of the population): $\sigma = \{p_1, \dots, p_i, \dots, p_n\}$, where p_i is the proportion of agents of selecting strategy s_i (type i) at a particular time; $p_i \geq 0$ and $\sum_i p_i = 1$

Evolutionarily stable strategies

- ***Fitness*** of an agent type *i* in a population = expected payoff from interaction with another member of population:

$$u_{i\sigma} = \sum_{j=1}^{n} u_{ij} p_j$$

- Mutant strategy $\tau = \{q_1, \ldots, q_i, \ldots, q_n\}$ ***invades*** *the population with* strategy σ ***at level*** ε (for small ε) if:
  - ε fraction of population uses τ
  - 1- ε fractionof population uses σ
  - The new state of the population is $\mu = (1-\varepsilon)\sigma + \varepsilon\tau$
- The payoff of a randomly chosen nonmutant is:

$$u_{\sigma\mu} = (1-\varepsilon)u_{\sigma\sigma} + \varepsilon u_{\sigma\tau}$$

- The payoff of a randomly chosen mutant is:

$$u_{\tau\mu} = (1-\varepsilon)u_{\tau\sigma} + \varepsilon u_{\tau\tau}$$

$$\text{where } u_{\sigma\sigma} = \sum_{i,j=1}^{n} p_i u_{ij} p_j,\ u_{\tau\sigma} = u_{\sigma\tau} = \sum_{i,j=1}^{n} q_i u_{ij} p_j,$$

$$u_{\tau\tau} = \sum_{i,j=1}^{n} q_i u_{ij} q_j$$

Evolutionarily stable strategies

- Strategy σ is ***evolutionarily stable*** if there is some number *y* such that:
  - When any other strategy τ invades σ at any level $\varepsilon < y$, the fitness of an agent playing σ is strictly greater than the fitness of an agent playing τ: $u_{\sigma\mu} > u_{\tau\mu}$

  That is $(1-\varepsilon)u_{\sigma\sigma} + \varepsilon u_{\sigma\tau} > (1-\varepsilon)u_{\tau\sigma} + \varepsilon u_{\tau\tau}$
- Given small ε, this is equivalent to requiring that either $u_{\sigma\sigma} > u_{\tau\sigma}$ or else both $u_{\sigma\sigma} = u_{\tau\sigma}$ and $u_{\sigma\tau} > u_{\tau\tau}$

Evolutionarily stable strategies

- Strategy σ is ***evolutionarily stable*** if there is some number *y* such that:
  - When any other strategy τ invades σ at any level $\varepsilon < y$, the fitness of an agent playing σ is strictly greater than the fitness of an agent playing τ: $u_{\sigma\mu} > u_{\tau\mu}$

  That is $(1-\varepsilon)u_{\sigma\sigma} + \varepsilon u_{\sigma\tau} > (1-\varepsilon)u_{\tau\sigma} + \varepsilon u_{\tau\tau}$

- Given small ε, this is equivalent to requiring that either $u_{\sigma\sigma} > u_{\tau\sigma}$ or else both $u_{\sigma\sigma} = u_{\tau\sigma}$ and $u_{\sigma\tau} > u_{\tau\tau}$

a mutant cannot do better against an existing agent than an existing agent can do against another existing agent

Evolutionarily stable strategies

- Strategy σ is ***evolutionarily stable*** if there is some number *y* such that:
  - When any other strategy τ invades σ at any level $\varepsilon < y$, the fitness of an agent playing σ is strictly greater than the fitness of an agent playing τ: $u_{\sigma\mu} > u_{\tau\mu}$

  That is $(1-\varepsilon)u_{\sigma\sigma} + \varepsilon u_{\sigma\tau} > (1-\varepsilon)u_{\tau\sigma} + \varepsilon u_{\tau\tau}$

- Given small ε, this is equivalent to requiring that either $u_{\sigma\sigma} > u_{\tau\sigma}$ or else both $u_{\sigma\sigma} = u_{\tau\sigma}$ and $u_{\sigma\tau} > u_{\tau\tau}$

But if a mutant does as well as an existing agent against another existing one, then an existing agent must do better against a mutant than a mutant does against another mutant.

Hawk–Dove game

- Two animals are fighting over a prize such as a piece of food.
- Each animal can choose between two behaviours:
  - an aggressive hawkish behaviour H, or
  - an gentle/peace dovish behaviour D.
- The prize is worth 6 to each of them.
- Fighting costs each player 5.
- When a hawk meets a dove he gets the prize without a fight, and hence the payoffs are 6 and 0, respectively.
- When two doves meet they split the prize without a fight, hence a payoff of 3 to each one.
- When two hawks meet a fight breaks out, costing each player 5 (or, equivalently, yielding −5). In addition, each player has a 50% chance of ending up with the prize,

	H	D
H	$-2, -2$	$6, 0$
D	$0, 6$	$3, 3$

Hawk–Dove game

- The game has a unique symmetric Nash equilibrium (σ, σ), where σ = (3/5, 2/5), and
- σ is also the unique ESS of the game.
- To confirm this, we need that $u(\sigma, \sigma) = u(\tau, \sigma)$ and $u(\sigma, \tau) > u(\tau, \tau)$

  For all $\tau \neq \sigma$,
  - The equality condition is true of any mixed strategy equilibrium with full support. Why???
  - The inequality also holds. To see this, consider

    $$f(\tau) = u(\sigma, \tau) - u(\tau, \tau)$$

  Expanding $f(\tau)$ we see that it is a quadratic equation with the (unique) maximum $\tau = \sigma$, proving our result

	H	D
H	$-2, -2$	$6, 0$
D	$0, 6$	$3, 3$

ESS Summary

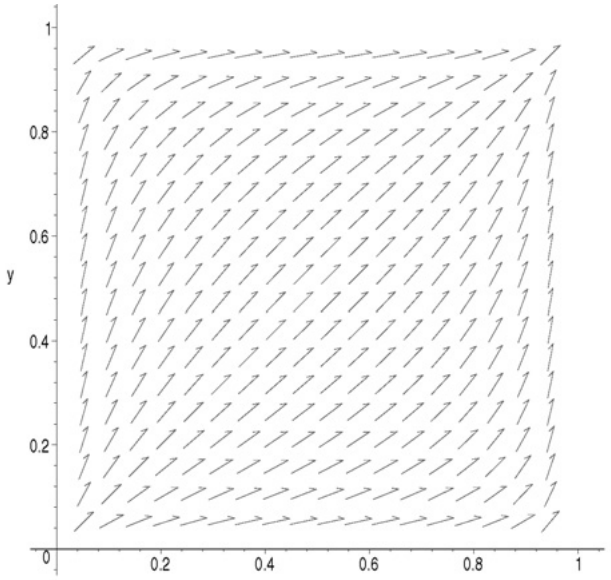
- Nash equilibrium
  - Rational players choosing mutual best responses to each other‘s strategy
  - Great demands on the ability to choose optimally and coordinate on strategies that are best responses to each other
- Evolutionarily stable strategies
  - No intelligence or coordination
  - Strategies hard-wired into players (genes)
  - Successful strategies produce more offspring
- Yet somehow they are almost the same!

Content

- IGA with Policy Prediction
- Gradient Ascent Optimization as Dynamical Systems
  - Dynamics in 2x2 Matrix Games
  - Equilibrium Points in Gradient Ascent Dynamics
  - Convergence Analysis via Lyapunov Functions
- Learning in Games
- Fictitious Play
- Smoothed Fictitious Play
- Rational Learning
- Evolutionary Game Theory
- Replicator Dynamics

Replicator Dynamics

- *Replicator dynamics* model a population undergoing frequent interactions.
- Focus on the symmetric, two-player case
  - A population of agents repeatedly play a two-player symmetric normal-form stage game against each other.
- It describes a population of agents playing such a game following:
  - At each point in time, each agent only plays a pure strategy.
  - Informally speaking, the model then pairs all agents and has them play each other, each obtaining some payoff. This payoff is called the agent's *fitness*.
  - Each agent now "reproduces" in a manner proportional to this fitness, and
  - the process repeats.
- The question is
  - whether the process converges to a fixed proportion of the various pure strategies within the population, and
  - if so to which fixed proportions.



Direction field plot of the Prisoner's dilemma game.

Tuyls, Karl, and Simon Parsons. "What evolutionary game theory tells us about multiagent learning." *Artificial Intelligence*171.7 (2007): 406-416.

Schuster, Peter, and Karl Sigmund. "Replicator dynamics." *Journal of theoretical biology* 100.3 (1983): 533-538.

Recall the stage game

- Consider a two-player normal form symmetric game:
  - both players have the set of pure strategies $S = \{s_1, \ldots, s_n\}$
  - $u_1(s_i, s_j)$: the payoffs of an agent playing s_i when playing with another agent using s_j.
  - Symmetric payoff: $u_1(s_i, s_j) = u_2(s_i, s_j) \equiv u_{i,j}$
  - Symmetric in strategy: agents cannot condition their play on whether they are player 1 or player 2
  - $\mathrm{U} = (u_{i,j})$: the matrix of the symmetric game

Replicator Dynamics

- Consider an evolutionary game where each player follows one of *n* pure strategies $s_i \in \{1, \dots, n\}$
- The play is repeated in periods t=1,2,…
- Let p_i^t be the fraction of players playing s_i in period t , and suppose the payoff to s_i is

$$u_{i\sigma}^t = \sum_{j=1}^{n} u_{ij} p_i^t$$

- For mathematical convenience, at a given time t , we index the strategies so that: $u_{1\sigma}^t \leq u_{2\sigma}^t \leq \cdots \leq u_{n\sigma}^t$

Replicator Dynamics

- Suppose in every time period *dt*, each agent with probability αdt > 0 learns the payoff to another randomly chosen agent and
- changes to the other's strategy if he perceives that the other's payoff is higher.
- However, information concerning the difference in the expected payoffs of the two strategies is imperfect, so the larger the difference in the payoffs, the more likely the agent is to perceive it, and change.
- Specifically, we assume that the probability $p_{i,j}^t$ that an agent using s_i will shift to s_j is given by

$$p_{i,j}^t = \begin{cases} \beta\left(u_{i\sigma}^t - u_{j\sigma}^t\right) & for \quad u_{i\sigma}^t \geq u_{j\sigma}^t \\ 0 & for \quad u_{j\sigma}^t \geq u_{i\sigma}^t \end{cases}$$

where β is sufficiently small that $p_{i,j}^t \leq 1$ holds for all i and j.

Replicator Dynamics

- The expected fraction of population using s_i in period t + dt is then given by

$$p_i^{t+dt} = p_i^t - \alpha dt \,(p_i^t \sum_{j=i+1}^{n} p_j^t \,\beta\left(u_{j\sigma}^t - u_{i\sigma}^t\right) + \sum_{j=1}^{i} p_j^t p_i^t \,\beta\left(u_{i\sigma}^t - u_{j\sigma}^t\right))$$

$$= p_i^t + \alpha dt \; p_i^t \sum_{j=1}^{n} p_j^t \,\beta\left(u_{i\sigma}^t - u_{j\sigma}^t\right)$$

$$= p_i^t + \alpha dt \; p_i^t \beta\left(u_{i\sigma}^t - \bar{u}_{.\sigma}^t\right)$$

where $\bar{u}_{.\sigma}^t$ is the average return for the whole population.

- Subtracting p_i^t from both sides, dividing by dt , and taking the limit as dt->0, we have

$$\dot{p}_i^t = \alpha \; \beta \; p_i^t\left(u_{i\sigma}^t - \bar{u}_{.\sigma}^t\right)$$

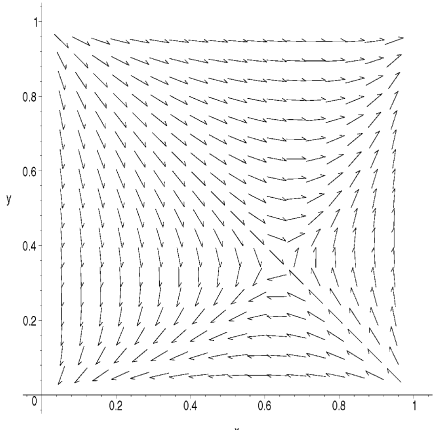
which is called the replicator dynamic.

As the constant $\alpha \; \beta$ merely changes the rate of adjustment to stationarity but leaves the stability properties and trajectories of the dynamical system unchanged, we often simply assume $\alpha \; \beta$=1

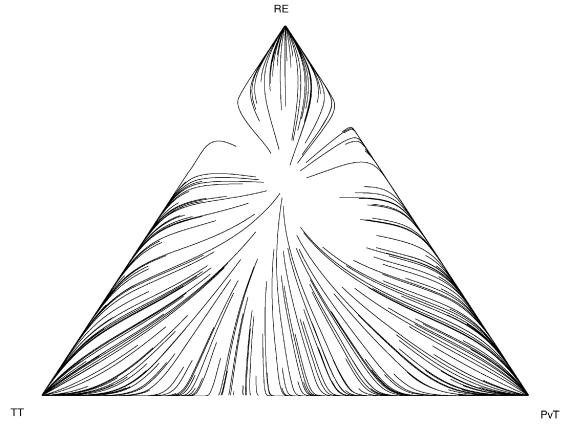
Gintis, Herbert. *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton university press, 2000.

Replicator Dynamics

- The system we have defined has a very intuitive quality.
  - If an action does better than the population average then the proportion of the population playing this action increases, and vice versa.
  - Note that even an action that is not a best response to the current population state can grow as a proportion of the population when its expected payoff is better than the population average.
- A straightforward interpretation is that it describes agents repeatedly interacting and replicating within a large population



Direction field plot of the battle of the sexes game.



Replicator dynamics direction field for CH with 10 agents.

Tuyls, Karl, and Simon Parsons. "What evolutionary game theory tells us about multiagent learning." *Artificial Intelligence*171.7 (2007): 406-416.

References

- Slides are partially based on *Ýmir Vigfússon's slides on the Structure of Networks*
- *Chapter 7, Networks, Crowds, and Markets: Reasoning about a Highly Connected World, By David Easley and Jon Kleinberg. Cambridge University Press, 2010.*
- Chapter 7, Shoham Y, Leyton-Brown K. Multiagent systems: Algorithmic, game-theoretic, and logical foundations[M]. Cambridge University Press, 2008.
- Gintis, Herbert. *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton university press, 2000.
- Fudenberg D, Levine DK. The theory of learning in games. MIT press; 1998
- Brown, George W. "Iterative solution of games by fictitious play." *Activity analysis of production and allocation* 13.1 (1951): 374-376.
- Fudenberg, Drew, and David K. Levine. "Consistency and cautious fictitious play." *Journal of Economic Dynamics and Control* 19.5-7 (1995): 1065-1089.
- Kalai, Ehud, and Ehud Lehrer. "Rational learning leads to Nash equilibrium." *Econometrica: Journal of the Econometric Society* (1993): 1019-1045.
- Tuyls, Karl, and Simon Parsons. "What evolutionary game theory tells us about multiagent learning." *Artificial Intelligence*171.7 (2007): 406-416.

References

- Bowling, Michael, and Manuela Veloso. "Multiagent learning using a variable learning rate." Artificial Intelligence 136.2 (2002): 215-250.
- "Method Of Lyapunov Functions". Math24, 2019, https://www.math24.net/method-lyapunov-functions/.
- Bloembergen, Daniël. Multi-agent learning dynamics. Maastricht University, 2015.
- Zhang, Chongjie, and Victor Lesser. "Multi-agent learning with policy prediction." Twenty-Fourth AAAI Conference on Artificial Intelligence. 2010.
- Bressan, Alberto. "Noncooperative differential games. a tutorial." Department of Mathematics, Penn State University (2010).
- Abdallah, Sherief, and Victor Lesser. "A multiagent reinforcement learning algorithm with non-linear dynamics." Journal of Artificial Intelligence Research 33 (2008): 521-549.
- Bhaya, Amit, Rodrigo Brandolt Sodré de Macedo, and Lucas Shiguemitsu Shigueoka. "Stability of the Nash equilibrium under gradient ascent learning algorithms in two-agent two-action games." 2013 IEEE International Conference on Control Applications (CCA). IEEE, 2013.