ΔΥΝΑΜΙΚΑ ΣΥΣΤΗΜΑΤΑ, ΕΞΕΛΙΚΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΚΑΙ ΜΑΘΗΣΗ ΣΤΗ ΘΕΩΡΙΑ ΠΑΙΓΝΙΩΝ

Παναγιώτης Μερτικόπουλος

⟨ Σεμινάριο Στατιστικής & Επιχ. Έρευνας | ΕΚΠΑ, Τμήμα Μαθηματικών | 4 Μαρτίου, 2022 ⟩

Outline

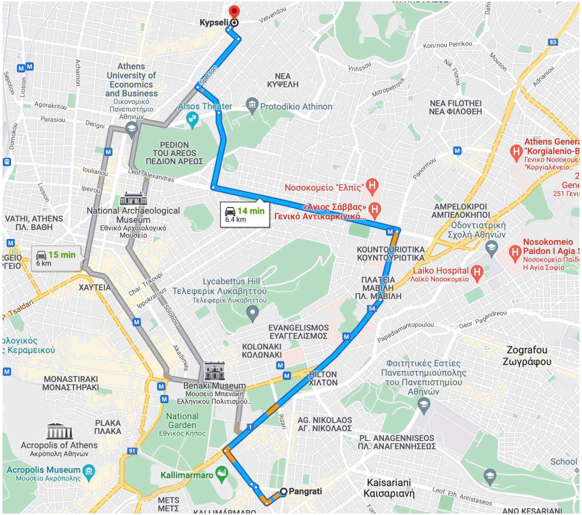
Traffic...

...how bad can it get?

Traffic...

...how bad can it get?

Game of roads



Athens at a glance

- $3,754,000$ people
- $937,000$ daily trips
- Up to 10^4 trips/min
- 1393 nodes
- 5429 edges
- $1,360,000$ O/D pairs
- $\approx 7 * 10^{18}$ paths

A very large game!

Online learning

A generic **online decision process:**

---

repeat

At each epoch t

Choose ***action*** — # single- / multi-player

Receive ***reward*** — # endogenous / exogenous

Get ***feedback*** (maybe) — # full info / oracle / payoff-based

until end

Online learning

A generic **online decision process:**

repeat
At each epoch t
  Choose ***action*** # single- / multi-player
  Receive ***reward*** # endogenous / exogenous
  Get ***feedback*** (maybe) # full info / oracle / payoff-based
until end

Defining elements

- **Time:** continuous or discrete?
- **Players:** continuous or finite?
- **Actions:** continuous or finite?
- **Reward mechanism:** endogenous or exogenous (determined by other players or by "Nature")?
- **Feedback:** observe other actions / other rewards / only received?

Online learning

A generic **online decision process:**

---

repeat

At each epoch t

Choose ***action*** # single- / multi-player

Receive ***reward*** # endogenous / exogenous

Get ***feedback*** (maybe) # full info / oracle / payoff-based

until end

Defining elements

- **Time:** continuous or discrete?
- **Players:** continuous or finite?
- **Actions:** ~~*continuous or*~~ finite
- **Reward mechanism:** endogenous ~~*or exogenous*~~ (determined by other players ~~*or by "Nature"*~~)
- **Feedback:** observe other actions / other rewards / only received?

Game-theoretic learning

- **Multiple agents**, individual objectives
- Payoffs determined by actions of **all** agents
- Agents receive payoffs, **adjust actions**, and the process repeats

Game-theoretic learning

- **Multiple agents**, individual objectives

[Select a route from home to work]

- Payoffs determined by actions of **all** agents

[Encounter other commuters on the road]

- Agents receive payoffs, **adjust actions**, and the process repeats

[Update road choice tomorrow]

Game-theoretic learning

- **Multiple agents**, individual objectives

[Select a route from home to work]

- Payoffs determined by actions of **all** agents

[Encounter other commuters on the road]

- Agents receive payoffs, **adjust actions**, and the process repeats

[Update road choice tomorrow]

Does learning lead to stable / rational outcomes?

Outline

1 Background

2 Preliminaries

3 Learning in continuous time

4 Learning in discrete time

Some basics

What's in a game?

A ***game in normal form*** is a collection of three basic elements:

1. A set of ***players*** $\mathcal{N}$
2. A set of ***actions*** (or ***pure strategies***) $\mathcal{A}_i$ per player $i \in \mathcal{N}$
3. An ensemble of ***payoff functions*** $u_i \colon \prod_j \mathcal{A}_j \to \mathbb{R}$ per player $i \in \mathcal{N}$

Some basics

What's in a game?

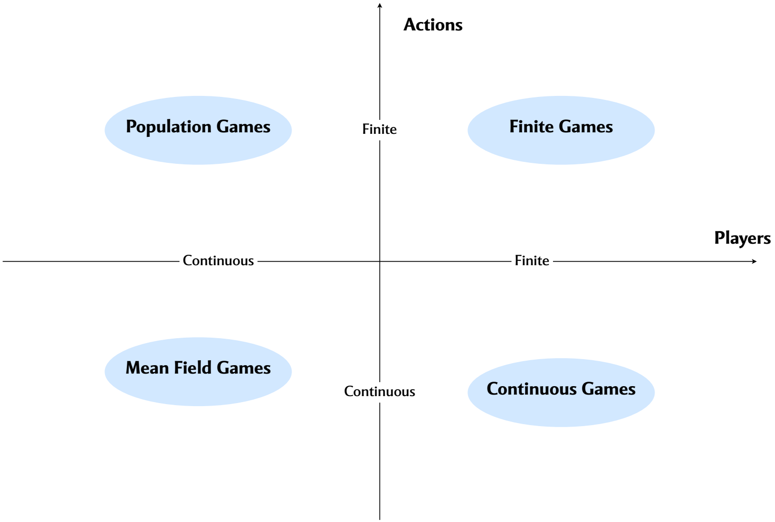
A ***game in normal form*** is a collection of three basic elements:

1. A set of ***players*** $\mathcal{N}$
2. A set of ***actions*** (or ***pure strategies***) $\mathcal{A}_i$ per player $i \in \mathcal{N}$
3. An ensemble of ***payoff functions*** $u_i \colon \prod_j \mathcal{A}_j \to \mathbb{R}$ per player $i \in \mathcal{N}$

Important:

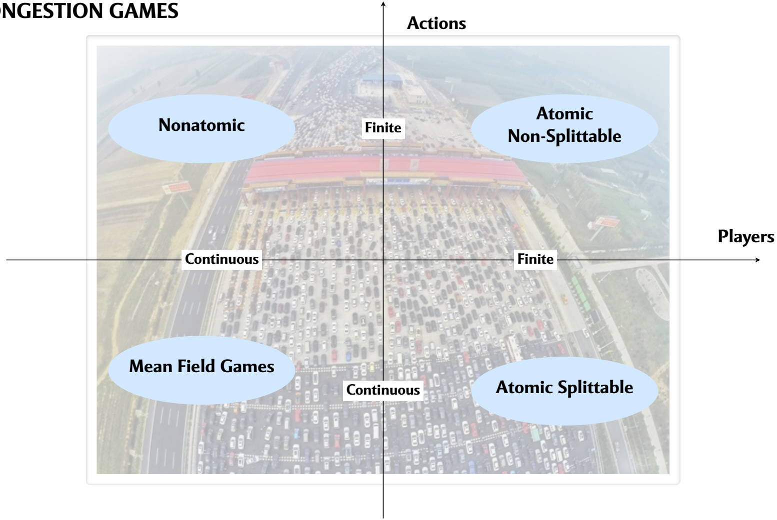
- Player set: atomic vs. nonatomic
- Action sets: finite vs. continuous; shared vs. individual; ...
- **NB:** do not mix game classes!

Taxonomy

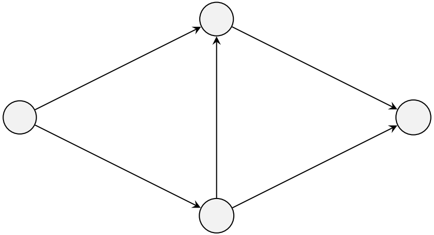

Actions
Population Games
Finite
Finite Games
Players
Continuous
Finite
Mean Field Games
Continuous
Continuous Games

Taxonomy

CONGESTION GAMES

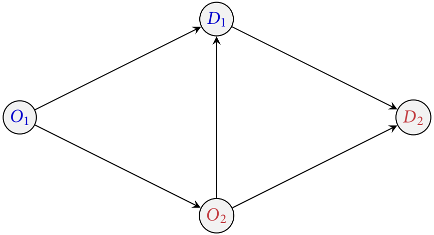


Nonatomic congestion games



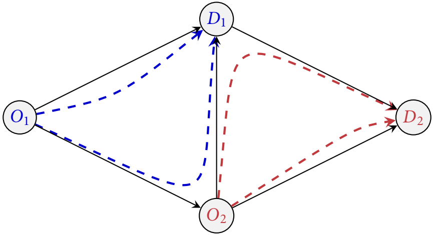
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Nonatomic congestion games



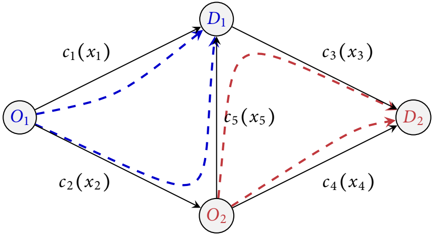
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs*** $i \in \mathcal{N}$: origin O_i sends m_i units of traffic to destination D_i [nonatomic, splittable]

Nonatomic congestion games



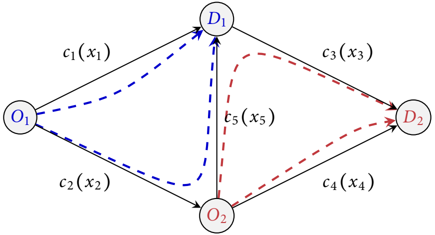
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** origin O_i sends m_i units of traffic to destination D_i [nonatomic, splittable]
- ***Paths $\mathcal{P}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$ [not necessarily *all* paths]
- ***Routing flow f_p:*** traffic along $p \in \mathcal{P} \equiv \bigcup_i \mathcal{P}_i$ generated by O/D pair owning p [congestion elements]

Nonatomic congestion games



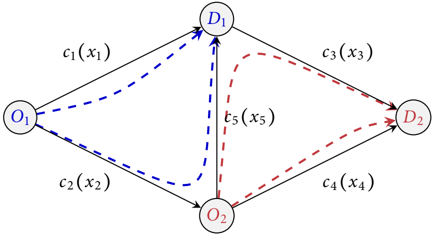
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs*** $i \in \mathcal{N}$: origin O_i sends m_i units of traffic to destination D_i [nonatomic, splittable]
- ***Paths*** $\mathcal{P}_i$: (sub)set of paths joining $O_i \rightsquigarrow D_i$ [not necessarily *all* paths]
- ***Routing flow*** f_p: traffic along $p \in \mathcal{P} \equiv \bigcup_i \mathcal{P}_i$ generated by O/D pair owning p [congestion elements]
- ***Load*** $x_e = \sum_{p \ni e} f_p$: total traffic along edge e [congestion mechanism]
- ***Edge cost function*** $c_e(x_e)$: cost along edge e when edge load is x_e [congestion cost]

Nonatomic congestion games



- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** origin O_i sends m_i units of traffic to destination D_i [nonatomic, splittable]
- ***Paths $\mathcal{P}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$ [not necessarily *all* paths]
- ***Routing flow f_p:*** traffic along $p \in \mathcal{P} \equiv \bigcup_i \mathcal{P}_i$ generated by O/D pair owning p [congestion elements]
- ***Load $x_e = \sum_{p \ni e} f_p$:*** total traffic along edge e [congestion mechanism]
- ***Edge cost function $c_e(x_e)$:*** cost along edge e when edge load is x_e [congestion cost]
- ***Path cost:*** $c_p(f) = \sum_{e \in p} c_e(x_e)$ [aggregate cost]
- **Nonatomic congestion game:** $\mathcal{C} = (\mathcal{G}, \mathcal{N}, \{m_i\}_{i \in \mathcal{N}}, \{\mathcal{P}_i\}_{i \in \mathcal{N}}, \{c_e\}_{e \in \mathcal{E}})$

Atomic congestion games



- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** origin O_i sends m_i units of traffic to destination D_i [atomic, non-splittable]
- ***Paths $\mathcal{P}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$ [not necessarily *all* paths]
- **Route choice $p_i \in \mathcal{P}_i$** $\rightsquigarrow$ congestion load of m_i units along each edge $e \in p_i$ [congestion elements]
- **Load $x_e = \sum_{p_i \ni e} m_i$:** total congestion load on edge e [congestion mechanism]
- ***Edge cost function $c_e(x_e)$:*** cost along edge e when edge load is x_e [congestion cost]
- ***Path cost:*** $c_p(f) = \sum_{e \in p} c_e(x_e)$ [aggregate cost]
- **Atomic congestion game:** $\mathcal{C} = (\mathcal{G}, \mathcal{N}, \{m_i\}_{i \in \mathcal{N}}, \{\mathcal{P}_i\}_{i \in \mathcal{N}}, \{c_e\}_{e \in \mathcal{E}})$

Finite games

Finite games:

[sometimes known as *(poly)matrix games*]

- Finite set of ***players*** $\mathcal{N} = \{1, \dots, N\}$
- Finite set of ***actions*** (or "***pure strategies***") $\mathcal{A}_i = \{1, \dots, m_i\}$ per player
- Action profile $a = (a_1, \dots, a_N) \in \mathcal{A} := \prod_i \mathcal{A}_i$
- Payoffs given by ***payoff functions*** $u_i \colon \mathcal{A} \to \mathbb{R}$

$$u_i(a) \equiv u_i(a_1, \dots, a_N) \equiv u_i(a_i; a_{-i})$$

- ***Payoff vector*** of player i:

$$v_i(a) = (u_i(a_i'; a_{-i}))_{a_i' \in \mathcal{A}_i}$$

- ***Notation:*** $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$

Mixed extensions

Mixed extension of a finite game:

- Given: finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$
- *Mixed strategy* of player i:

$$x_i = (x_{ia})_{a \in \mathcal{A}_i} \in \Delta(\mathcal{A}_i) =: \mathcal{X}_i$$

- *Mixed payoff* of player i

$$u_i(x) = \mathbb{E}_{a \sim x}\, u_i(a) = \sum_{a_1 \in \mathcal{A}_1} \dots \sum_{a_N \in \mathcal{A}_N} x_{1,a_1} \cdots x_{N,a_N}\, u_i(a_1, \dots, a_N)$$

- *Payoff vector* of player i:

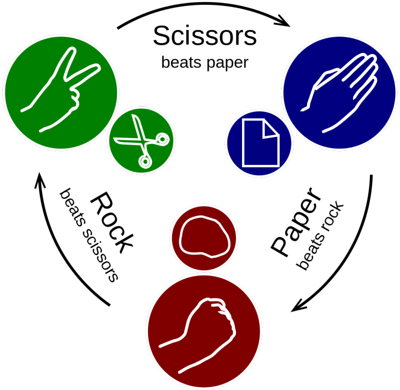
$$v_i(a) = (u_i(a'_i; a_{-i}))_{a'_i \in \mathcal{A}_i}$$

- *Notation:* $\bar{\Gamma} \equiv \Delta(\Gamma)$

Toy example: Rock-Paper-Scissors

Playing with mixed strategies:

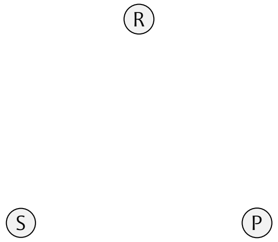
- Players: $\mathcal{N} = \{1, 2\}$



Toy example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$

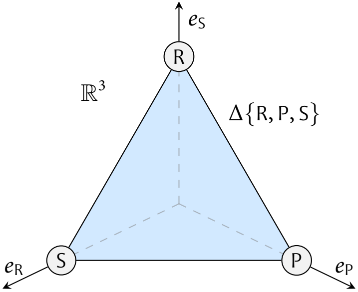


$$M = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

Toy example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Mixed strategy space: $\mathcal{X}_i = \Delta\{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$

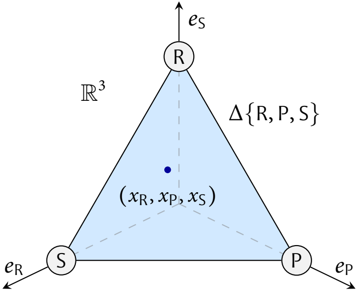


$$M = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

Toy example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Mixed strategy space: $\mathcal{X}_i = \Delta\{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Choose mixed strategy $x_i \in \mathcal{X}_i$

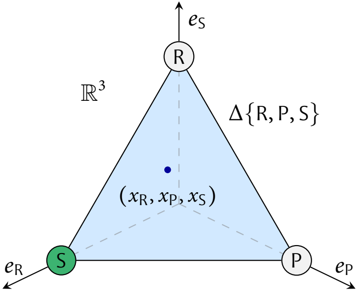


$$M = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

Toy example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Mixed strategy space: $\mathcal{X}_i = \Delta\{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Choose mixed strategy $x_i \in \mathcal{X}_i$
- Choose action $a_i \sim x_i$



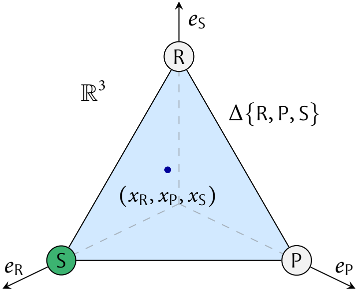
$$M = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

Toy example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Mixed strategy space: $\mathcal{X}_i = \Delta\{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Choose mixed strategy $x_i \in \mathcal{X}_i$
- Choose action $a_i \sim x_i$
- Mixed strategy payoffs:

$$u_1(x_1, x_2) = x_1^\top M x_2$$
$$u_2(x_1, x_2) = -u_1(x_1, x_2)$$



$$M = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

Single-population games

- ***Population of players:*** $\mathcal{I} = [0,1]$ [endowed with Lebesgue measure μ]
- Common set of actions $\mathcal{A} = \{1, \dots, m\}$
- **Strategy profile:** measurable function $\chi : \mathcal{N} \to \mathcal{A}$ [measurable assignment of players to actions]
- ***Population state*** $x \coloneqq \chi \sharp \mu \equiv \mu \circ \chi^{-1}$, i.e., [viewed as element of $\mathcal{X} \coloneqq \Delta(\mathcal{A})$]

$$x_a = \mu(\chi^{-1}(a)) = \text{mass of players playing } a \in \mathcal{A}$$

- Payoffs given by ***payoff functions*** $v_a : \mathcal{X} \to \mathbb{R}$ [Players are ***anonymous***]

$$v_a(x) = \text{payoff to } a\text{-strategists when the population is at state } x \in \mathcal{X}$$

- ***Mean population payoff:*** $u(x) = \sum_a x_a v_a(x)$

Single-population games

- ***Population of players:*** $\mathcal{I} = [0,1]$ [endowed with Lebesgue measure μ]
- Common set of actions $\mathcal{A} = \{1, \dots, m\}$
- **Strategy profile:** measurable function $\chi : \mathcal{N} \to \mathcal{A}$ [measurable assignment of players to actions]
- ***Population state*** $x \coloneqq \chi \sharp \mu \equiv \mu \circ \chi^{-1}$, i.e., [viewed as element of $\mathcal{X} \coloneqq \Delta(\mathcal{A})$]

$$x_a = \mu(\chi^{-1}(a)) = \text{mass of players playing } a \in \mathcal{A}$$

- Payoffs given by ***payoff functions*** $v_a : \mathcal{X} \to \mathbb{R}$ [Players are ***anonymous***]

$$v_a(x) = \text{payoff to } a\text{-strategists when the population is at state } x \in \mathcal{X}$$

- ***Mean population payoff:*** $u(x) = \sum_a x_a v_a(x)$

Example (Symmetric / Single-population random matching)

- ***Given:*** symmetric $m \times m$ payoff matrix M
- Players drawn randomly from population at state x to play M
- Mean payoff to a-strategists: $v_a(x) = \sum_{a' \in \mathcal{A}} M_{aa'} x_{a'} = (Mx)_a$

Multi-population games

- ***Multiple populations:*** $\mathcal{I} = [0,1] \times \cdots \times [0,1]$ [endowed with Lebesgue measure μ]
- Population-specific action sets $\mathcal{A}_i$, $i = 1, \dots, N$
- ***Population state*** $x \in \mathcal{X} := \prod_i \Delta(\mathcal{A}_i)$

$$x_{ia_i} = \text{mass of players of population } i \text{ playing } a_i \in \mathcal{A}_i$$

- Payoffs given by ***payoff functions*** $v_{ia_i} \colon \mathcal{X} \to \mathbb{R}$

$$v_{ia_i}(x) = \text{payoff to } a_i\text{-strategists when the population is at state } x \in \mathcal{X}$$

- ***Mean population payoff:*** $u_i(x) = \sum_{a_i \in \mathcal{A}_i} x_{ia_i} v_{ia_i}(x)$

Multi-population games

- ***Multiple populations:*** $\mathcal{I} = [0,1] \times \cdots \times [0,1]$ [endowed with Lebesgue measure μ]
- Population-specific action sets $\mathcal{A}_i$, $i = 1, \dots, N$
- ***Population state*** $x \in \mathcal{X} := \prod_i \Delta(\mathcal{A}_i)$

$$x_{ia_i} = \text{mass of players of population } i \text{ playing } a_i \in \mathcal{A}_i$$

- Payoffs given by ***payoff functions*** $v_{ia_i} \colon \mathcal{X} \to \mathbb{R}$

$$v_{ia_i}(x) = \text{payoff to } a_i\text{-strategists when the population is at state } x \in \mathcal{X}$$

- ***Mean population payoff:*** $u_i(x) = \sum_{a_i \in \mathcal{A}_i} x_{ia_i} v_{ia_i}(x)$

Example (Asymmetric / Multi-population random matching)

- ***Given:*** finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$
- N players drawn randomly from each population to play Γ
- Mean payoff to a_i-strategists in the i-th population: $v_{ia_i}(x) = u_i(a_i; x_{-i})$

Mix'n'match

☞ **Symmetric Random matching ≠ Mixed extension**

[Population matched against itself $\Longrightarrow$ ***symmetric interactions***]

Mix'n'match

☞ **Symmetric Random matching $\neq$ Mixed extension**

[Population matched against itself $\Longrightarrow$ ***symmetric interactions***]

☞ ***Asymmetric* random matching = Mixed Extension**

[Populations matched against each other $\Longrightarrow$ ***asymmetric interactions***]

Mix'n'match

☞ **Symmetric Random matching ≠ Mixed extension**

[Population matched against itself $\implies$ ***symmetric interactions***]

☞ ***Asymmetric*** **random matching = Mixed Extension**

[Populations matched against each other $\implies$ ***asymmetric interactions***]

☞ **Multi-population games $\supsetneq$ Mixed Extensions**

[Nonatomic congestion games, ...]

Nash equilibrium

Equilibrium principle (Nash, 1950, 1951)

"No player has an incentive to deviate from their chosen strategy if other players don't"

Nash equilibrium

Equilibrium principle (Nash, 1950, 1951)

"No player has an incentive to deviate from their chosen strategy if other players don't"

- **In finite games** (mixed extension formulation):

$$u_i(x_i^*; x_{-i}^*) \geq u_i(x_i; x_{-i}^*) \quad \text{for all } x_i \in \mathcal{X}_i,\, i \in \mathcal{N}$$

- **In population games:**

$$v_{ia_i}(x^*) \geq v_{ia_i'}(x^*) \quad \text{whenever } a_i \in \operatorname{supp}(x^*)$$

Nash equilibrium

Equilibrium principle (Nash, 1950, 1951)

"No player has an incentive to deviate from their chosen strategy if other players don't"

- **In finite games** (mixed extension formulation):

$$u_i(x_i^*; x_{-i}^*) \geq u_i(x_i; x_{-i}^*) \quad \text{for all } x_i \in \mathcal{X}_i, i \in \mathcal{N}$$

- **In population games:**

$$v_{ia_i}(x^*) \geq v_{ia_i'}(x^*) \quad \text{whenever } a_i \in \operatorname{supp}(x^*)$$

Variational formulation (Stampacchia, 1964)

$$\langle v(x^*), x - x^* \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}$$

where $v(x) = (v_1(x), \dots, v_N(x))$ is the ***payoff field*** of the game

[Geometric interpretation: $v(x^*)$ is outward-pointing]

Learning, evolution and dynamics

What is "learning" in games?

Learning, evolution and dynamics

What is “learning” in games?

The basic process:

- Players choose strategies and receive corresponding payoffs
- Depending on outcome and information revealed, they choose new strategies and they play again
- Rinse, repeat

Learning, evolution and dynamics

What is "learning" in games?

The basic process:

- Players choose strategies and receive corresponding payoffs
- Depending on outcome and information revealed, they choose new strategies and they play again
- Rinse, repeat

The basic questions:

- *How do populations evolve over time?* [Population biology]
- *How do people learn in a game?* [Economics]
- *What algorithms should we use to learn in a game?* [Computer science]
- ***Given a dynamical system on $\mathcal{X}$, what is its long-term behavior?*** [Mathematics]

Outline

Age the First (1970's-1990's): Population Biology

- Strategies are ***phenotypes*** in a given species

$$z_a = \text{absolute population mass of type } a \in \mathcal{A}$$

$$z = \sum_a z_a = \text{absolute population mass}$$

Age the First (1970's-1990's): Population Biology

- Strategies are ***phenotypes*** in a given species

$$z_a = \text{absolute population mass of type } a \in \mathcal{A}$$

$$z = \sum_a z_a = \text{absolute population mass}$$

- Utilities measure ***fecundity / reproductive fitness:***

$$v_a = \text{per capita growth rate of type } a$$

- Population evolution:

$$\dot{z}_a = z_a v_a$$

Age the First (1970's-1990's): Population Biology

- Strategies are ***phenotypes*** in a given species

$$z_a = \text{absolute population mass of type } a \in \mathcal{A}$$
$$z = \sum_a z_a = \text{absolute population mass}$$

- Utilities measure ***fecundity / reproductive fitness:***

$$v_a = \text{per capita growth rate of type } a$$

- Population evolution:

$$\dot{z}_a = z_a v_a$$

- Evolution of population shares ($x_a = z_a/z$):

$$\dot{x}_a = \frac{d}{dt}\frac{z_a}{z} = \frac{\dot{z}_a z - z_a \sum_{a'} \dot{z}_{a'}}{z^2} = \frac{z_a}{z} v_a - \frac{z_a}{z} \sum_{a'} \frac{z_{a'}}{z} v_{a'}$$

Age the First (1970's-1990's): Population Biology

- Strategies are ***phenotypes*** in a given species

$$z_a = \text{absolute population mass of type } a \in \mathcal{A}$$
$$z = \sum_a z_a = \text{absolute population mass}$$

- Utilities measure ***fecundity / reproductive fitness:***

$$v_a = \text{per capita growth rate of type } a$$

- Population evolution:

$$\dot{z}_a = z_a v_a$$

- Evolution of population shares ($x_a = z_a/z$):

$$\dot{x}_a = \frac{d}{dt}\frac{z_a}{z} = \frac{\dot{z}_a z - z_a \sum_{a'} \dot{z}_{a'}}{z^2} = \frac{z_a}{z} v_a - \frac{z_a}{z} \sum_{a'} \frac{z_{a'}}{z} v_{a'}$$

Replicator dynamics (Taylor & Jonker, 1978)

$$\dot{x}_a = x_a [v_a(x) - u(x)] \tag{RD}$$

Age the Second (1990's-2010's): Economics

- Agents receive **revision opportunities** to switch strategies

$$\rho_{aa'}(x) = \text{conditional switch rate from } a \text{ to } a'$$

[**NB:** dropping player index for simplicity]

Age the Second (1990's-2010's): Economics

- Agents receive **revision opportunities** to switch strategies

$$\rho_{aa'}(x) = \text{conditional switch rate from } a \text{ to } a'$$

[**NB:** dropping player index for simplicity]

- **Pairwise proportional imitation:**

$$\rho_{aa'}(x) = x_{a'}[v_{a'}(x) - v_a(x)]_+$$

[Imitate with probability proportional to excess payoff (Helbing, 1992; Schlag, 1998)]

Age the Second (1990's-2010's): Economics

- Agents receive **revision opportunities** to switch strategies

$$\rho_{aa'}(x) = \text{conditional switch rate from } a \text{ to } a'$$

[**NB:** dropping player index for simplicity]

- **Pairwise proportional imitation:**

$$\rho_{aa'}(x) = x_{a'}[v_{a'}(x) - v_a(x)]_+$$

[Imitate with probability proportional to excess payoff (Helbing, 1992; Schlag, 1998)]

- Inflow/outflow:

$$\textit{Incoming toward } a = \sum_{a'} \text{mass}(a' \rightsquigarrow a) = \sum_{a' \in \mathcal{A}} x_{a'} \rho_{a'a}(x)$$
$$\textit{Outgoing from } a = \sum_{a'} \text{mass}(a \rightsquigarrow a') = x_a \sum_{a' \in \mathcal{A}} \rho_{aa'}(x)$$

Age the Second (1990's-2010's): Economics

- Agents receive **revision opportunities** to switch strategies

$$\rho_{aa'}(x) = \text{conditional switch rate from } a \text{ to } a'$$

[**NB:** dropping player index for simplicity]

- **Pairwise proportional imitation:**

$$\rho_{aa'}(x) = x_{a'}[v_{a'}(x) - v_a(x)]_+$$

[Imitate with probability proportional to excess payoff (Helbing, 1992; Schlag, 1998)]

- Inflow/outflow:

$$\textit{Incoming toward } a = \sum_{a'} \text{mass}(a' \rightsquigarrow a) = \sum_{a' \in \mathcal{A}} x_{a'} \rho_{a'a}(x)$$
$$\textit{Outgoing from } a = \sum_{a'} \text{mass}(a \rightsquigarrow a') = x_a \sum_{a' \in \mathcal{A}} \rho_{aa'}(x)$$

- Detailed balance:

$$\dot{x}_a = \text{inflow}_a(x) - \text{outflow}_a(x) = \cdots = x_a[v_a(x) - u(x)] \tag{RD}$$

Age the Third (2000's-present): Computer Science

Evolution of mixed strategies in a **finite game:**

- Agents record cumulative payoff of each strategy

$$y_a(t) = \int_0^t v_a(\tau)\,d\tau$$

$\Longrightarrow$ ***propensity*** of choosing a strategy [Auer et al., 1995; Freund & Schapire, 1999; Littlestone & Warmuth, 1994]

Age the Third (2000's-present): Computer Science

Evolution of mixed strategies in a **finite game:**

- Agents record cumulative payoff of each strategy

$$y_a(t) = \int_0^t v_a(\tau)\, d\tau$$

$\Longrightarrow$ ***propensity*** of choosing a strategy [Auer et al., 1995; Freund & Schapire, 1999; Littlestone & Warmuth, 1994]

- Choice probabilities $\rightsquigarrow$ exponentially proportional to propensity scores

$$x_a(t) \propto \exp(y_a(t))$$

Age the Third (2000's-present): Computer Science

Evolution of mixed strategies in a **finite game:**

- Agents record cumulative payoff of each strategy

$$y_a(t) = \int_0^t v_a(\tau)\, d\tau$$

$\Longrightarrow$ ***propensity*** of choosing a strategy [Auer et al., 1995; Freund & Schapire, 1999; Littlestone & Warmuth, 1994]

- Choice probabilities $\rightsquigarrow$ exponentially proportional to propensity scores

$$x_a(t) = \frac{\exp(y_a(t))}{\sum_{a'} \exp(y_{a'}(t))}$$

Age the Third (2000's-present): Computer Science

Evolution of mixed strategies in a **finite game:**

- Agents record cumulative payoff of each strategy

$$y_a(t) = \int_0^t v_a(\tau)\, d\tau$$

$\Longrightarrow$ ***propensity*** of choosing a strategy [Auer et al., 1995; Freund & Schapire, 1999; Littlestone & Warmuth, 1994]

- Choice probabilities $\rightsquigarrow$ exponentially proportional to propensity scores

$$x_a(t) = \frac{\exp(y_a(t))}{\sum_{a'} \exp(y_{a'}(t))}$$

- Evolution of mixed strategies [Hofbauer et al., 2009; Rustichini, 1999]

$$\dot{x}_a = \cdots = x_a[v_a(x) - u(x)] \tag{RD}$$

Basic properties

Multi-player replicator dynamics

$$\dot{x}_{ia_i} = x_{a_i}\left[v_{ia_i}(x) - u_i(x)\right] \tag{RD}$$

[**NB:** focus on multi-population version from now on]

Basic properties

Multi-player replicator dynamics

$$\dot{x}_{ia_i} = x_{a_i}\left[v_{ia_i}(x) - u_i(x)\right] \tag{RD}$$

[**NB:** focus on multi-population version from now on]

Structural properties

[Hofbauer & Sigmund, 1998; Weibull, 1995]

- **Well-posed:** every initial condition $x \in \mathcal{X}$ admits unique solution trajectory $x(t)$ that exists for all time

  [Assuming u_i is Lipschitz]

- **Consistent:** $x(t) \in \mathcal{X}$ for all $t \geq 0$

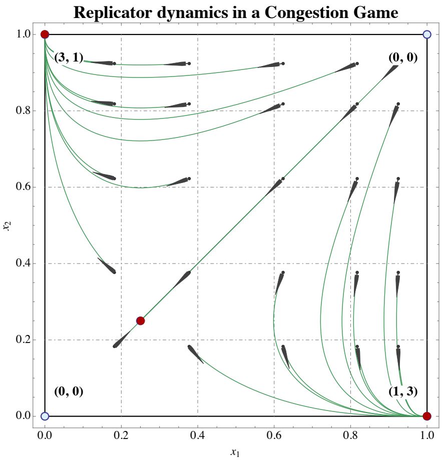
  [Assuming $x(0) \in \mathcal{X}$]

- **Faces are forward invariant** ("strategies breed true"):

$$x_{ia_i}(0) > 0 \iff x_{ia_i}(t) > 0 \quad \text{for all } t \geq 0$$
$$x_{ia_i}(0) = 0 \iff x_{ia_i}(t) = 0 \quad \text{for all } t \geq 0$$

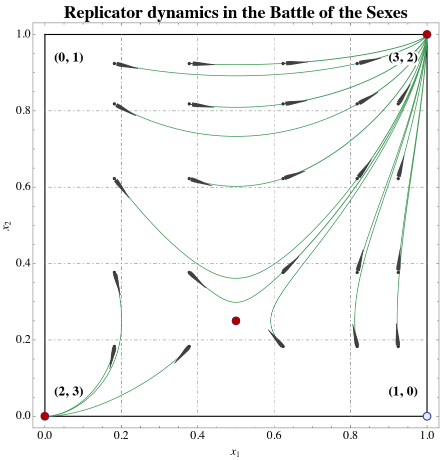
Phase portraits

What do the dynamics look like?



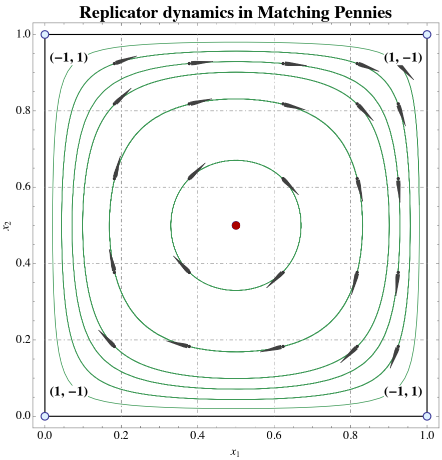
Phase portraits

What do the dynamics look like?



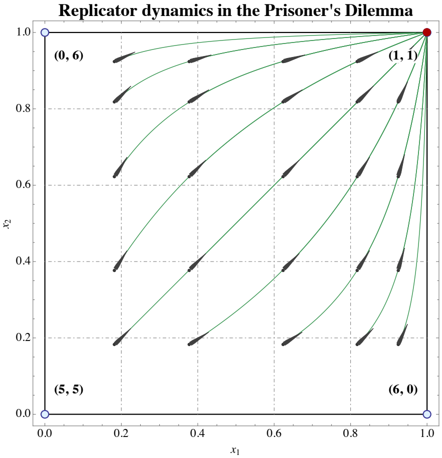
Phase portraits

What do the dynamics look like?



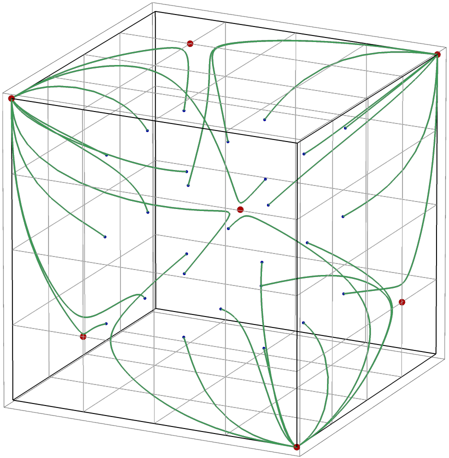
Phase portraits

What do the dynamics look like?



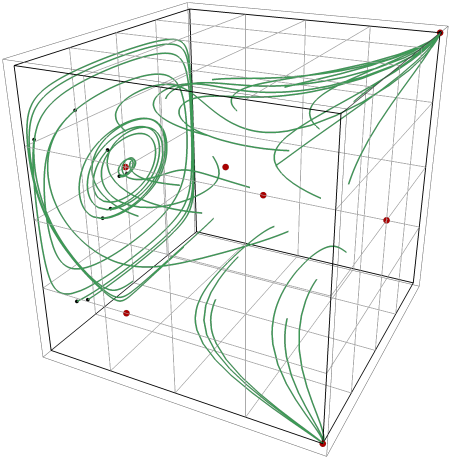
Phase portraits

What do the dynamics look like?



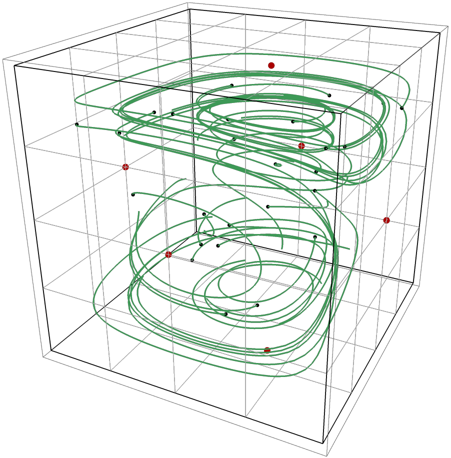
Phase portraits

What do the dynamics look like?



Phase portraits

What do the dynamics look like?



Rationality analysis

Are game-theoretic solution concepts consistent with the players' dynamics?

- Are Nash equilibria stationary?
- Are they ***stable?*** Are they ***attracting?***
- Do the replicator dynamics always converge?
- What other behaviors can we observe?
- …

Stationarity of equilibria

Equilibrium: $v_{ia_i}(x^*) \geq v_{ia_i'}(x^*)$ for all $a_i, a_i' \in \mathcal{A}_i$ with $x^*_{ia_i} > 0$

- Supported strategies have equal payoffs:

$$v_{ia_i}(x^*) = v_{ia_i'}(x^*) \quad \text{for all } a_i, a_i' \in \operatorname{supp}(x_i^*)$$

- Mean payoff equal to equilibrium payoff:

$$u_i(x^*) = v_{ia_i}(x^*) \quad \text{for all } a_i \in \operatorname{supp}(x_i^*)$$

- Replicator field vanishes at Nash equilibria:

$$x^*_{ia_i}[v_{ia_i}(x^*) - u_i(x^*)] = 0 \quad \text{for all } a_i \in \mathcal{A}_i$$

Stationarity of equilibria

Equilibrium: $v_{ia_i}(x^*) \geq v_{ia'_i}(x^*)$ for all $a_i, a'_i \in \mathcal{A}_i$ with $x^*_{ia_i} > 0$

- Supported strategies have equal payoffs:

$$v_{ia_i}(x^*) = v_{ia'_i}(x^*) \quad \text{for all } a_i, a'_i \in \operatorname{supp}(x^*_i)$$

- Mean payoff equal to equilibrium payoff:

$$u_i(x^*) = v_{ia_i}(x^*) \quad \text{for all } a_i \in \operatorname{supp}(x^*_i)$$

- Replicator field vanishes at Nash equilibria:

$$x^*_{ia_i}[v_{ia_i}(x^*) - u_i(x^*)] = 0 \quad \text{for all } a_i \in \mathcal{A}_i$$

Proposition (Stationarity of Nash equilibria)

Let $x(t)$ be a solution orbit of (RD). *Then:*

$$x(0) \text{ is a Nash equilibrium} \implies x(t) = x(0) \text{ for all } t \geq 0$$

Stationarity of equilibria

Equilibrium: $v_{ia_i}(x^*) \geq v_{ia'_i}(x^*)$ for all $a_i, a'_i \in \mathcal{A}_i$ with $x^*_{ia_i} > 0$

- Supported strategies have equal payoffs:

$$v_{ia_i}(x^*) = v_{ia'_i}(x^*) \quad \text{for all } a_i, a'_i \in \operatorname{supp}(x^*_i)$$

- Mean payoff equal to equilibrium payoff:

$$u_i(x^*) = v_{ia_i}(x^*) \quad \text{for all } a_i \in \operatorname{supp}(x^*_i)$$

- Replicator field vanishes at Nash equilibria:

$$x^*_{ia_i}[v_{ia_i}(x^*) - u_i(x^*)] = 0 \quad \text{for all } a_i \in \mathcal{A}_i$$

Proposition (Stationarity of Nash equilibria)

Let $x(t)$ be a solution orbit of (RD). *Then:*

$$x(0) \text{ is a Nash equilibrium} \implies x(t) = x(0) \text{ for all } t \geq 0$$

✗ **The converse does not hold!**

[See previous portraits]

Stability

Are all stationary points created equal?

Definition (Lyapunov stability)

x^* is ***(Lyapunov) stable*** if, for every neighborhood $\mathcal{U}$ of x^* in $\mathcal{X}$, there exists a neighborhood $\mathcal{U}'$ of x^* such that

$$x(0) \in \mathcal{U}' \implies x(t) \in \mathcal{U} \quad \text{for all } t \geq 0$$

[Trajectories that start close to x^* remain close for all time]

Stability

Are all stationary points created equal?

Definition (Lyapunov stability)

x^* is ***(Lyapunov) stable*** if, for every neighborhood $\mathcal{U}$ of x^* in $\mathcal{X}$, there exists a neighborhood $\mathcal{U}'$ of x^* such that

$$x(0) \in \mathcal{U}' \implies x(t) \in \mathcal{U} \quad \text{for all } t \geq 0$$

[Trajectories that start close to x^* remain close for all time]

Proposition (Folk)

Suppose that x^* is Lyapunov stable under (RD). *Then x^* is a Nash equilibrium.*

Asymptotic stability

Are all equilibria created equal?

Definition

- x^* is ***attracting*** if $\lim_{t\to\infty} x(t) = x^*$ whenever $x(0)$ is close enough to x^*
- x^* is ***asymptotically stable*** if it is stable and attracting

Asymptotic stability

Are all equilibria created equal?

Definition

- x^* is ***attracting*** if $\lim_{t\to\infty} x(t) = x^*$ whenever $x(0)$ is close enough to x^*
- x^* is ***asymptotically stable*** if it is stable and attracting

Proposition (Folk)

Strict Nash equilibria are asymptotically stable under (RD).

The "folk theorem" of evolutionary game theory

Theorem ("folk"; Hofbauer & Sigmund, 2003)

Let Γ be a finite game. Then, under (RD), we have:

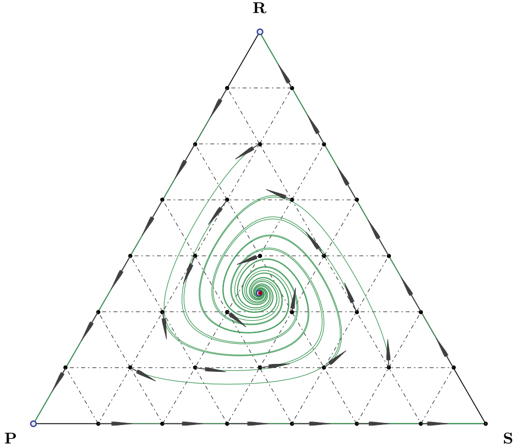
1. x^* *is a Nash equilibrium* $\implies$ x^* *is stationary*
2. x^* *is the limit of an interior trajectory* $\implies$ x^* *is a Nash equilibrium*
3. x^* *is stable* $\implies$ x^* *is a Nash equilibrium*
4. x^* *is asymptotically stable* $\iff$ x^* *is a strict Nash equilibrium*

Notes:

- ✍ Concerns **multi-population** replicator dynamics
- ✗ **Converse to (1), (2) and (3) does not hold!**
- ✍ Symmetric version: all true except $\implies$ in (4)

Single-population: different ball game

The replicator dynamics in “good” RPS (win > loss):



Convergence in potential games

Potential games (Sandholm, 2001)

$$v_{ia_i} = -\frac{\partial \Phi}{\partial x_{ia_i}} \quad \text{for some } \textit{potential function } \Phi\colon \mathcal{X} \to \mathbb{R}$$

NASC (Poincaré's lemma):

$$\text{potential} \iff \frac{\partial v_{ia_i}}{\partial x_{ia'_i}} = \frac{\partial v_{ia'_i}}{\partial x_{ia_i}}$$

Convergence in potential games

Potential games (Sandholm, 2001)

$$v_{ia_i} = -\frac{\partial \Phi}{\partial x_{ia_i}} \quad \text{for some } \textit{potential function } \Phi\colon \mathcal{X} \to \mathbb{R}$$

NASC (Poincaré's lemma):

$$\text{potential} \iff \frac{\partial v_{ia_i}}{\partial x_{ia_i'}} = \frac{\partial v_{ia_i'}}{\partial x_{ia_i}}$$

Positive correlation / Lyapunov property:

$$\frac{d\Phi}{dt} \leq 0 \quad \text{under (RD)}$$

Convergence in potential games

Potential games (Sandholm, 2001)

$$v_{ia_i} = -\frac{\partial \Phi}{\partial x_{ia_i}} \quad \text{for some } \textbf{\textit{potential function}}\ \Phi\colon \mathcal{X} \to \mathbb{R}$$

NASC (Poincaré's lemma):

$$\text{potential} \iff \frac{\partial v_{ia_i}}{\partial x_{ia_i'}} = \frac{\partial v_{ia_i'}}{\partial x_{ia_i}}$$

Positive correlation / Lyapunov property:

$$\frac{d\Phi}{dt} \leq 0 \quad \text{under (RD)}$$

Theorem (Sandholm, 2001)

- *In potential games,* (RD) *converges to its set of stationary points*
- *In random matching potential games, interior trajectories of* (RD) *converge to Nash equilibrium*

Non-convergence in zero-sum games

The landscape is very different in zero-sum games:

Non-convergence in zero-sum games

The landscape is very different in zero-sum games:

x^* is full-support equilibrium $\implies$ (RD) admits a **constant of motion**

KL divergence: $D_{\mathrm{KL}}(x^*, x) = \sum_i \sum_{a_i} x^*_{ia_i} \log \dfrac{x^*_{ia_i}}{x_{ia_i}}$

Non-convergence in zero-sum games

The landscape is very different in zero-sum games:

$$x^* \text{ is full-support equilibrium} \implies \text{(RD) admits a } \textbf{constant of motion}$$

KL divergence: $D_{\mathrm{KL}}(x^*, x) = \sum_i \sum_{a_i} x^*_{ia_i} \log \frac{x^*_{ia_i}}{x_{ia_i}}$

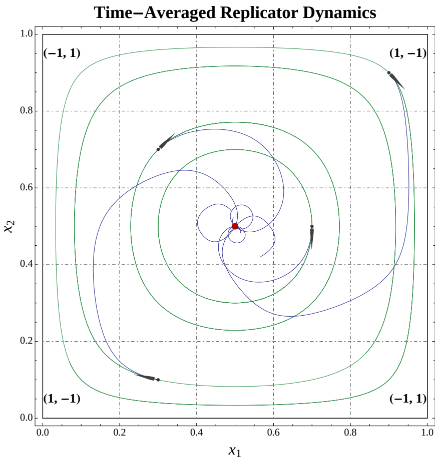
Theorem (Hofbauer et al., 2009)

Assume a bilinear zero-sum game admits an interior equilibrium. Then:

- *Interior trajectories of* (RD) ***do not converge*** *(unless stationary)*
- *Time-averages* $\bar{x}(t) = t^{-1} \int_0^t x(\tau)\, d\tau$ ***converge to Nash equilibrium***

Convergence of time-averages

The replicator dynamics in a game of Matching Pennies



Poincaré recurrence in zero-sum games

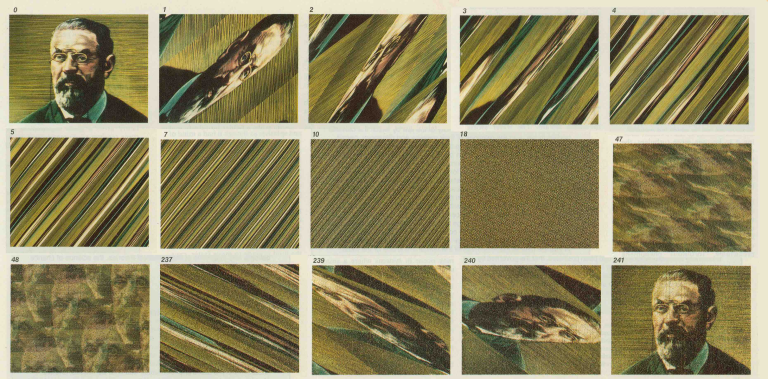
Definition (Poincaré)

A dynamical system is **Poincaré recurrent** if almost all solution trajectories return ***arbitrarily close*** to their starting point ***infinitely many times***

Poincaré recurrence in zero-sum games

Definition (Poincaré)

A dynamical system is **Poincaré recurrent** if almost all solution trajectories return ***arbitrarily close*** to their starting point ***infinitely many times***



Poincaré recurrence in zero-sum games

Proposition (Coucheney et al., 2015)

The dynamics (RD) *are volume-preserving under the Shahshahani metric* $g_{aa'}(x) = \delta_{aa'}/x_a$ on $\operatorname{ri}\mathcal{X}$.

Volume preservation $\implies$ **no concentration** $\implies$ no convergence ✓

Poincaré recurrence in zero-sum games

Proposition (Coucheney et al., 2015)

The dynamics (RD) *are volume-preserving under the Shahshahani metric* $g_{aa'}(x) = \delta_{aa'}/x_a$ on $\operatorname{ri}\mathcal{X}$.

Volume preservation $\implies$ **no concentration** $\implies$ no convergence ✓

...but the Shahshahani metric becomes **singular** at the boundary of $\mathcal{X}$ ✗

Poincaré recurrence in zero-sum games

Proposition (Coucheney et al., 2015)

The dynamics (RD) *are volume-preserving under the Shahshahani metric* $g_{aa'}(x) = \delta_{aa'}/x_a$ on $\operatorname{ri}\mathcal{X}$.

Volume preservation $\implies$ **no concentration** $\implies$ no convergence ✓

...but the Shahshahani metric becomes **singular** at the boundary of $\mathcal{X}$ ✗

Theorem (M et al., 2018)

(RD) *is Poincaré recurrent in all bilinear zero-sum games with a full-support equilibrium*

Poincaré recurrence in zero-sum games

Proposition (Coucheney et al., 2015)

The dynamics (RD) *are volume-preserving under the Shahshahani metric* $g_{aa'}(x) = \delta_{aa'}/x_a$ on $\operatorname{ri}\mathcal{X}$.

Volume preservation $\implies$ **no concentration** $\implies$ no convergence ✓

...but the Shahshahani metric becomes singular at the boundary of $\mathcal{X}$ ✗

Theorem (M et al., 2018)

(RD) *is Poincaré recurrent in all bilinear zero-sum games with a full-support equilibrium*

Theorem (Coucheney et al., 2015; M & Sandholm, 2016; Flokas et al., 2020)

Any attractor of (RD) *contains a pure strategy.*

Follow the regularized leader

Are the nice propeties of (RD) a "fluke"?

Follow the regularized leader

Are the nice propeties of (RD) a "fluke"?

- The logit map $\Lambda(y) = (\exp(y_a))_{a\in\mathcal{A}} \big/ \sum_a \exp(y_a)$ approximates the "***leader***" (best response map)

$$y \mapsto \arg\max_{x\in\mathcal{X}} \langle y, x\rangle$$

Follow the regularized leader

Are the nice propeties of (RD) a "fluke"?

- The logit map $\Lambda(y) = (\exp(y_a))_{a \in \mathcal{A}} / \sum_a \exp(y_a)$ approximates the "***leader***" (best response map)

$$y \mapsto \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$$

where $h(x) = \sum_{a \in \mathcal{A}} x_a \log x_a$ is the (negative) entropy of $x \in \mathcal{X}$

Follow the regularized leader

Are the nice propeties of (RD) a "fluke"?

- The logit map $\Lambda(y) = (\exp(y_a))_{a\in\mathcal{A}} / \sum_a \exp(y_a)$ approximates the "***leader***" (best response map)

$$y \mapsto \arg\max_{x\in\mathcal{X}} \{\langle y, x\rangle - h(x)\}$$

  where $h(x) = \sum_{a\in\mathcal{A}} x_a \log x_a$ is the (negative) entropy of $x \in \mathcal{X}$

- **Regularized best responses**

$$Q(y) = \arg\max_{x\in\mathcal{X}} \{\langle y, x\rangle - h(x)\}$$

  where $h\colon \mathcal{X} \to \mathbb{R}$ is a (strictly) convex **regularizer function**

Follow the regularized leader

Are the nice propeties of (RD) a "fluke"?

- The logit map $\Lambda(y) = (\exp(y_a))_{a\in\mathcal{A}} / \sum_a \exp(y_a)$ approximates the "***leader***" (best response map)

$$y \mapsto \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$$

  where $h(x) = \sum_{a\in\mathcal{A}} x_a \log x_a$ is the (negative) entropy of $x \in \mathcal{X}$

- **Regularized best responses**

$$Q(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$$

  where $h\colon \mathcal{X} \to \mathbb{R}$ is a (strictly) convex **regularizer function**

Follow the regularized leader (FTRL)

$$\begin{aligned} \dot{y}_t &= v_t \\ x_t &= Q(y_t) \end{aligned} \tag{FTRL}$$

The projection dynamics

Example: Quadratic (Euclidean) regularization

$$h(x) = \frac{1}{2} \sum_a x_a^2$$

The projection dynamics

Example: Quadratic (Euclidean) regularization

$$h(x) = \frac{1}{2} \sum_a x_a^2$$

Choice map $\rightsquigarrow$ closest point projection:

$$\Pi(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - (1/2)\|x\|_2^2\} = \arg\min_{x \in \mathcal{X}} \|y - x\|$$

The projection dynamics

Example: Quadratic (Euclidean) regularization

$$h(x) = \frac{1}{2} \sum_a x_a^2$$

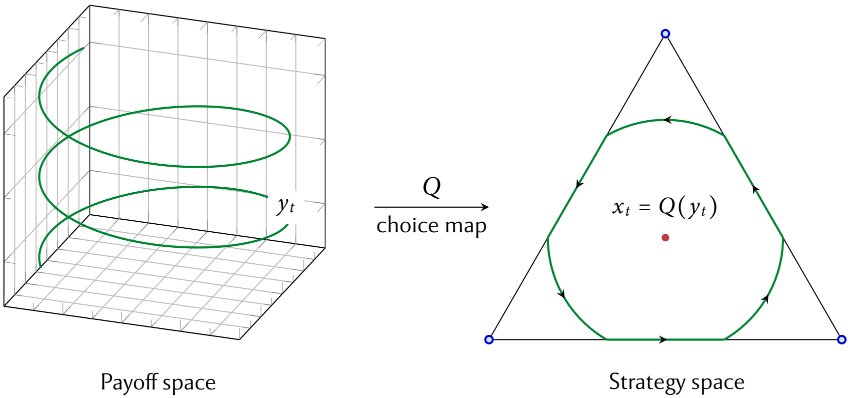
Choice map $\rightsquigarrow$ closest point projection:

$$\Pi(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - (1/2)\|x\|_2^2\} = \arg\min_{x \in \mathcal{X}} \|y - x\|$$

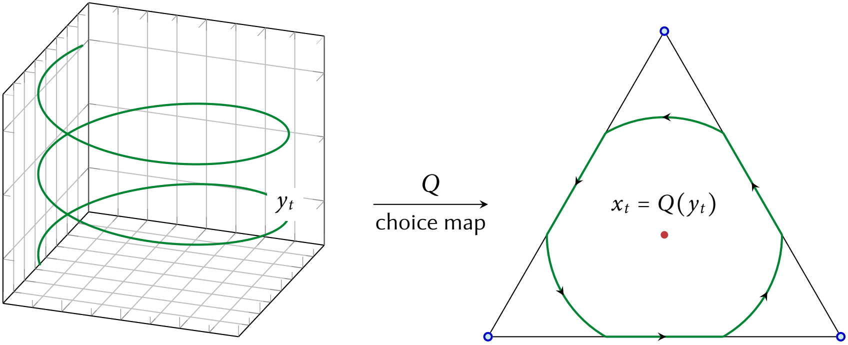
Projection dynamics [M & Sandholm, 2016]

$$\begin{aligned} \dot{y}_t &= v_t \\ x_t &= \Pi(y_t) \end{aligned} \tag{PL}$$

In and out of the boundary



In and out of the boundary



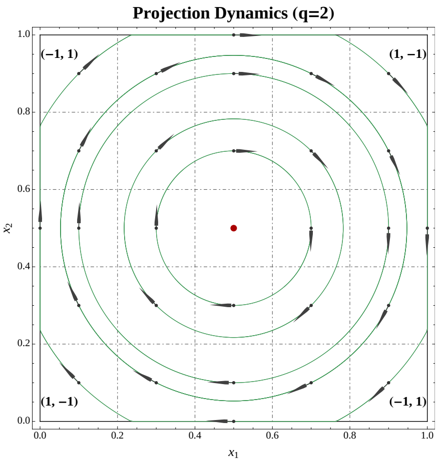
Payoff space

Strategy space

Key difference with replicator: faces no longer forward invariant

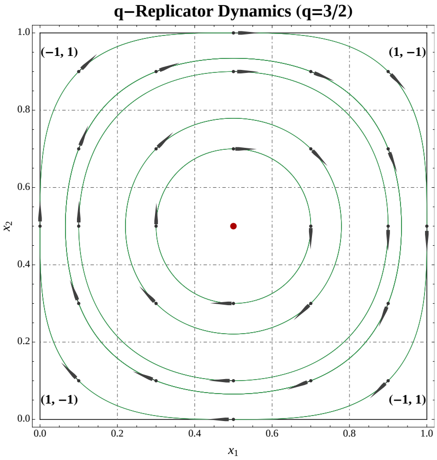
Portraits and examples

The Tsallis-Havrda -Charvát kernel: $h(x) = [q(1-q)]^{-1} \sum_a (x_a - x_a^q)$



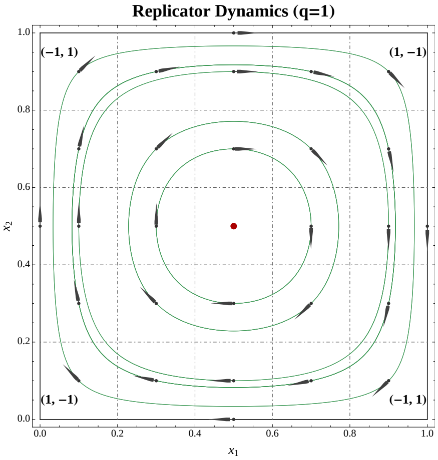
Portraits and examples

The Tsallis-Havrda -Charvát kernel: $h(x) = [q(1-q)]^{-1} \sum_a (x_a - x_a^q)$



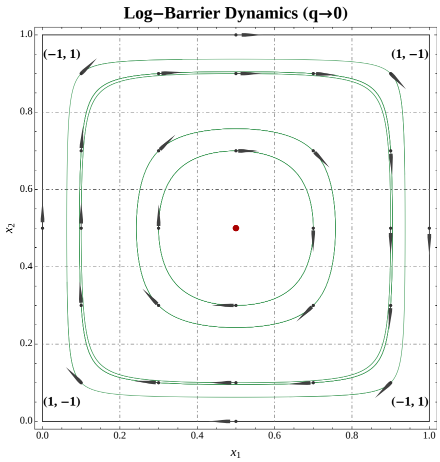
Portraits and examples

The Tsallis-Havrda -Charvát kernel: $h(x) = [q(1-q)]^{-1} \sum_a (x_a - x_a^q)$



Portraits and examples

The Tsallis-Havrda -Charvát kernel: $h(x) = [q(1-q)]^{-1} \sum_a (x_a - x_a^q)$



Rational behavior under FTRL

Do the rationality properties of (RD) extend to (FTRL)?

Rational behavior under FTRL

Do the rationality properties of (RD) extend to (FTRL)?

Theorem (Coucheney et al., 2015; M & Sandholm, 2016; Flokas et al., 2020)

Let Γ be a finite game. Then, under (FTRL)*, we have:*

1. x^* *is a Nash equilibrium* $\implies$ x^* *is stationary*
2. x^* *is the limit of an interior trajectory* $\implies$ x^* *is a Nash equilibrium*
3. x^* *is stable* $\implies$ x^* *is a Nash equilibrium*
4. x^* *is asymptotically stable* $\iff$ x^* *is a strict Nash equilibrium*

Outline

The model

Require: finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ # atomic setting

repeat

At each epoch $n = 1, 2, \dots$ **do simultaneously** for all players $i \in \mathcal{N}$ # discrete time

Choose ***mixed strategy*** $X_{i,n} \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # mixed extension

Choose ***action*** $a_{i,n} \sim X_{i,n}$ # random action selection

Observe mixed payoff vector $v_i(X_{i,n}; X_{-i,n})$ # feedback phase

until end

Defining elements

- ***Time:*** $n = 1, 2, \dots$
- ***Players:*** finite
- ***Actions:*** finite
- ***Mixing:*** yes
- ***Feedback:*** mixed payoff vectors

The model

Require: finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ # atomic setting

repeat

At each epoch $n = 1, 2, \dots$ **do simultaneously** for all players $i \in \mathcal{N}$ # discrete time

Choose ***mixed strategy*** $X_{i,n} \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # mixed extension

Choose ***action*** $a_{i,n} \sim X_{i,n}$ # random action selection

Observe pure payoff vector $v_i(a_{i,n}; a_{-i,n})$ # feedback phase

until end

Defining elements

- ***Time:*** $n = 1, 2, \dots$
- ***Players:*** finite
- ***Actions:*** finite
- ***Mixing:*** yes
- ***Feedback:*** pure payoff vectors

The model

Require: finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ # atomic setting

repeat

At each epoch $n = 1, 2, \dots$ **do simultaneously** for all players $i \in \mathcal{N}$ # discrete time

Choose ***mixed strategy*** $X_{i,n} \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # mixed extension

Choose ***action*** $a_{i,n} \sim X_{i,n}$ # random action selection

Observe realized payoff $u_i(a_{i,n}; a_{-i,n})$ # feedback phase

until end

Defining elements

- ***Time:*** $n = 1, 2, \dots$
- ***Players:*** finite
- ***Actions:*** finite
- ***Mixing:*** yes
- ***Feedback:*** realized payoffs

The feedback process

Different types of feedback (from best to worst):

- ***Mixed payoff vectors***: $v_i(X_{i,n}; X_{-i,n})$
- ***Pure payoff vectors***: $v_i(a_{i,n}; a_{-i,n})$
- ***Bandit / Payoff-based:*** $u_{i,n}(a_{i,n}; a_{-i,n})$

The feedback process

Different types of feedback (from best to worst):

- ***Mixed payoff vectors***: $v_i(X_{i,n}; X_{-i,n})$
- ***Pure payoff vectors***: $v_i(a_{i,n}; a_{-i,n})$
- ***Bandit / Payoff-based:*** $u_{i,n}(a_{i,n}; a_{-i,n})$

Features:

- ***Vector*** (mixed / pure payoff vecs) vs. ***Scalar*** (bandit)
- ***Deterministic*** (mixed payoff vecs) vs. ***Stochastic*** (pure payoff vecs, bandit)

NB1: Randomness defined relative to ***history of play*** $\mathcal{F}_n := \mathcal{F}(X_1, \dots, X_n)$

NB2: Other feedback models also possible (noisy observations,...)

From payoffs to payoff vectors

How to estimate the payoff $u_i(a_i; a_{-i,n})$ of an unplayed action $a_i \neq a_{i,n}$?

From payoffs to payoff vectors

How to estimate the payoff $u_i(a_i; a_{-i,n})$ of an unplayed action $a_i \neq a_{i,n}$?

Definition (Importance weighted estimation)

The ***importance weighted estimator*** of a vector $v \in \mathbb{R}^{\mathcal{A}}$ given a mixed strategy $x \in \Delta(\mathcal{A})$ is defined as

$$\hat{v}_a = \frac{\mathbb{1}_a}{x_a} v_a = \begin{cases} v_a / x_a & \text{if } a \text{ is drawn } (a = \hat{a}) \\ 0 & \text{otherwise} \quad (a \neq \hat{a}) \end{cases} \tag{IWE}$$

From payoffs to payoff vectors

How to estimate the payoff $u_i(a_i; a_{-i,n})$ of an unplayed action $a_i \neq a_{i,n}$?

Definition (Importance weighted estimation)

The ***importance weighted estimator*** of a vector $v \in \mathbb{R}^{\mathcal{A}}$ given a mixed strategy $x \in \Delta(\mathcal{A})$ is defined as

$$\hat{v}_a = \frac{\mathbb{1}_a}{x_a} v_a = \begin{cases} v_a / x_a & \text{if } a \text{ is drawn } (a = \hat{a}) \\ 0 & \text{otherwise} \quad (a \neq \hat{a}) \end{cases} \tag{IWE}$$

Statistical properties of (IWE)

- ***Unbiased:***

$$\mathbb{E}[\hat{v}_a] = v_a$$

- ***Second moment:***

$$\mathbb{E}[\hat{v}_a^2] = \frac{v_a^2}{x_a}$$

The oracle model

Definition (Black-box oracle)

A ***stochastic first-order oracle*** of $v(X_n)$ is a random vector of the form

$$\hat{v}_n = v(X_n) + U_n + b_n$$

where U_n is **zero-mean** and $b_n = \mathbb{E}[\hat{v}_n \mid \mathcal{F}_n] - v(X_n)$ is the **bias** of $\hat{v}_n$.

The oracle model

Definition (Black-box oracle)

A ***stochastic first-order oracle*** of $v(X_n)$ is a random vector of the form

$$\hat{v}_n = v(X_n) + U_n + b_n$$

where U_n is **zero-mean** and $b_n = \mathbb{E}[\hat{v}_n \mid \mathcal{F}_n] - v(X_n)$ is the **bias** of $\hat{v}_n$.

Examples

- Mixed payoff vectors: $\hat{v}_{i,n} = v_i(X_{i,n}; X_{-i,n})$ [noise $U_n = 0$; bias $b_n = 0$]
- Pure payoff vectors: $\hat{v}_{i,n} = v_i(a_{i,n}; a_{-i,n})$ [noise $U_n = \mathcal{O}(1)$; bias $b_n = 0$]
- Payoff-based: $\hat{v}_{i,n} = \dfrac{u_i(a_{i,n}; a_{-i,n})}{\mathbb{P}(a_{i,n} = a_i)} e_{a_{i,n}}$ [noise $U_n = \mathcal{O}(1/\min_{a_i} x_{ia_i,n})$; bias $b_n = 0$]

Follow the regularized leader in discrete time

The FTRL template

$$
\begin{aligned}
Y_{i,n+1} &= Y_{i,n} + \gamma_n \hat{v}_{i,n} \\
X_{i,n+1} &= Q_i(Y_{i,n+1}) \equiv \arg\max_{x_i \in \mathcal{X}_i} \{ \langle Y_{n+1}, x \rangle - h_i(x_i) \}
\end{aligned}
\tag{FTRL}
$$

[Algorithm due to Shalev-Shwartz, 2011; Shalev-Shwartz & Singer, 2006]

- $\gamma_n > 0$ is the method's step-size [To be specialized later]
- $\hat{v}_{i,n}$ is an stochastic first-order oracle (SFO) model for $v_i(x_n)$ [To be specialized later]
- Every player's ***regularizer*** $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is continuous on $\mathcal{X}_i$, differentiable on $\operatorname{ri} \mathcal{X}_i$, and strongly convex on $\mathcal{X}_i$

$$
h_i(x_i') \geq h_i(x_i) + \langle \nabla h_i(x_i), x_i' - x_i \rangle + (K_i/2) \| x_i' - x_i \|^2
$$

Examples

Example 1: Ridge regularization

- **Regularizer:**

$$h(x) = \frac{1}{2}\|x\|^2$$

- **Algorithm:**

$$Y_{n+1} = Y_n + \gamma_n \hat{v}_n \qquad X_{n+1} = \Pi_{\mathcal{X}}(Y_{n+1})$$

Examples

Example 1: Ridge regularization

- **Regularizer:**

$$h(x) = \tfrac{1}{2}\|x\|^2$$

- **Algorithm:**

$$Y_{n+1} = Y_n + \gamma_n \hat{v}_n \qquad X_{n+1} = \Pi_X(Y_{n+1})$$

Example 2: Entropic regularization

- **Regularizer:**

$$h(x) = \sum_{a\in\mathcal{A}} x_a \log x_a$$

- **Algorithm:**

$$Y_{n+1} = Y_n + \gamma_n \hat{v}_n \qquad X_{n+1} = \Lambda(Y_{n+1})$$

Exponential weights redux

Algorithm Exponential weights in discrete time (ExpWeight)

Require: finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$; stochastic first-order oracle $\hat{v}$

Initialize: $Y_i \in \mathbb{R}^{\mathcal{A}_i}$, $i = 1, \dots, N$

for all $n = 1, 2, \dots$ all players $i \in \mathcal{N}$ **do simultaneously**

  set $X_{i,n} \propto \exp(Y_{i,n})$ # mixed strategy

  play $a_{i,n} \sim X_{i,n}$ # choose action

  get $\hat{v}_{i,n} \in \mathbb{R}^{\mathcal{A}_i}$ # receive feedback

  set $Y_{i,n+1} \leftarrow Y_{i,n} + \gamma_n \hat{v}_{i,n}$ # update scores

end for

Basic idea:

- Score actions by aggregating payoff vector estimates provided by oracle
- Choose actions with probability exponentially proportional to their scores
- Rinse / repeat

Model 1: ExpWeight with mixed payoff vector observations

If players observe ***mixed payoff vectors:***

$$\hat{v}_{i,n} = v_i(X_{i,n}; X_{-i,n})$$

Oracle features:

- ***Deterministic:*** no randomness!
- ***Bias:*** $B_n = 0$
- ***Variance:*** $\sigma_n = 0$
- ***Second moment:*** $M_n = \mathcal{O}(1)$

Model 2: ExpWeight with pure payoff vector observations

If players observe ***pure payoff vectors:***

$$\hat{v}_{i,n} = v_i(a_{i,n}; a_{-i,n})$$

Oracle features:

- ***Stochastic:*** random action selection
- ***Bias:*** $B_n = 0$
- ***Variance:*** $\sigma_n = \mathcal{O}(1)$
- ***Second moment:*** $M_n = \mathcal{O}(1)$

NB: this algorithm is known as as **Hedge** [Auer et al., 1995, 2002,]

Model 3: ExpWeight with bandit feedback

If players observe ***realized payoffs only:***

$$\hat{v}_{i,n} = \frac{u_i(a_{i,n}; a_{-i,n})}{\mathbb{P}(a_{i,n} = a_i)} e_{a_{i,n}}$$

Oracle features:

- ***Stochastic:*** random action selection
- ***Bias:*** $B_n = 0$
- ***Variance:*** $\sigma_n = \mathcal{O}(1/X_{ia_i,n})$
- ***Second moment:*** $M_n = \mathcal{O}(1/X_{ia_i,n})$

NB: this algorithm is known as as **EXP3** [Auer et al., 1995, 2002,]

Model 4: ExpWeight with bandit feedback

If players observe ***realized payoffs only:***

$$\hat{v}_{i,n} = \frac{u_i(a_{i,n}; a_{-i,n})}{\mathbb{P}(a_{i,n} = a_i)} \, e_{a_{i,n}}$$

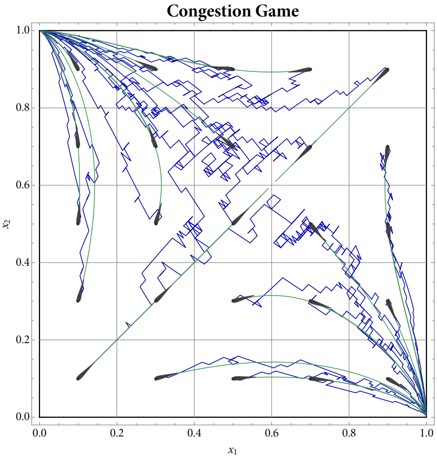
Oracle features:

- ***Stochastic:*** random action selection
- ***Explicit exploration:*** draw $a_{i,n} \sim X_{i,n}$ with prob. $1 - \delta_n$, otherwise uniformly
- ***Bias:*** $B_n = \mathcal{O}(\delta_n)$
- ***Variance:*** $\sigma_n = \mathcal{O}(1/\delta_n^2)$
- ***Second moment:*** $M_n = \mathcal{O}(1/\delta_n^2)$

NB: this algorithm is known as as **EXP3 with Explicit Exploration** [Lattimore & Szepesvári, 2020; Shalev-Shwartz, 2011]

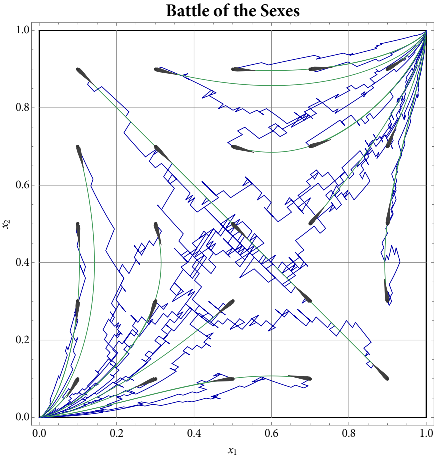
Visualization

What does the sequence of play look like?



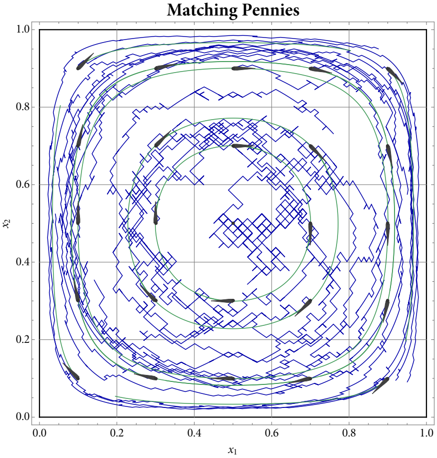
Visualization

What does the sequence of play look like?



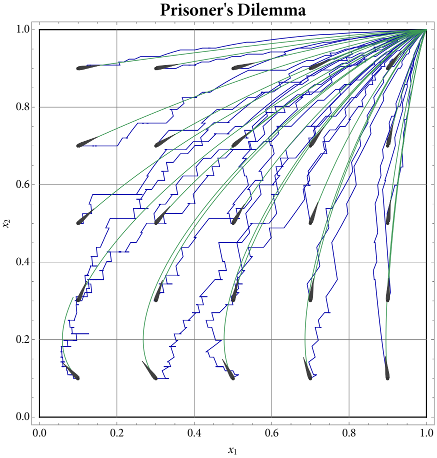
Visualization

What does the sequence of play look like?



Visualization

What does the sequence of play look like?



Notions of stability

Definition (Stochastic stability)

$x^* \in \mathcal{X}$ is ***stochastically stable*** under X_n if, for every confidence level $\delta > 0$ and every neighborhood $\mathcal{U}$ of x^*, there exists a neighborhood $\mathcal{U}_1$ of x^* such that

$$\mathbb{P}(X_n \in \mathcal{U} \text{ for all } n = 1, 2, \dots \mid X_1 \in \mathcal{U}_1) \geq 1 - \delta$$

[**Intuition:** with high probability, if X_n starts near x^*, it remains nearby]

Notions of stability

Definition (Stochastic stability)

$x^* \in \mathcal{X}$ is ***stochastically stable*** under X_n if, for every confidence level $\delta > 0$ and every neighborhood $\mathcal{U}$ of x^*, there exists a neighborhood $\mathcal{U}_1$ of x^* such that

$$\mathbb{P}(X_n \in \mathcal{U} \text{ for all } n = 1, 2, \dots \mid X_1 \in \mathcal{U}_1) \geq 1 - \delta$$

[**Intuition:** with high probability, if X_n starts near x^*, it remains nearby]

Definition (Stochastic asymptotic stability)

- $x^* \in \mathcal{X}$ is ***attracting*** if, for every confidence level $\delta > 0$, there exists a neighborhood $\mathcal{U}_1$ of x^* such that

$$\mathbb{P}(X_n \to x^* \; as \; n \to \infty \mid X_1 \in \mathcal{U}_1) \geq 1 - \delta$$

- $x^* \in \mathcal{X}$ is ***stochastically asymptotically stable*** if it is stochastically stable and attracting.

[**Intuition:** with high probability, if X_n starts near x^* then, it remains nearby and eventually converges to x^*]

The behavior of regularized learning in games

Theorem

➥ ***Assume:*** *all players run (FTRL) with step-size γ_n and oracle parameters b_n (bias) and U_n (noise) such that:*

(A1) $\gamma_n > 0$ and $\sum_n \gamma_n = \infty$

(A2) $b_n \to 0$

(A3) $\mathbb{E}[\|U_n\|^q] \leq \sigma_n^q$ for some $q > 2$

(A4) $\sum_{k=1}^n \gamma_k^{1+q/2} \sigma_k^q / [\sum_{k=1}^n \gamma_k]^{1+\alpha q}$ is summmable for some $\alpha \in (0,1)$

The behavior of regularized learning in games

Theorem

➥ **Assume:** *all players run* (FTRL) *with step-size* γ_n *and oracle parameters* b_n *(bias) and* U_n *(noise) such that:*

(A1) $\gamma_n = \gamma/n^p$ *for some* $p \in [0,1]$

(A2) $b_n = \mathcal{O}(1/n^b)$ *for some* $b > 0$

(A3) $\mathbb{E}[\|U_n\|^q] = \mathcal{O}(1/n^r)$ *for some* $q > 2$, $r < 1/2$

The behavior of regularized learning in games

Theorem

➥ ***Assume:*** *all players run* (FTRL) *with step-size γ_n and oracle parameters b_n (bias) and U_n (noise) such that:*

(A1) $\gamma_n = \gamma/n^p$ *for some* $p \in [0,1]$

(A2) $b_n = \mathcal{O}(1/n^b)$ *for some* $b > 0$

(A3) $\mathbb{E}[\|U_n\|^q] = \mathcal{O}(1/n^r)$ *for some* $q > 2$, $r < 1/2$

✍ ***Then:*** *the sequence X_n generated by* (FTRL) *enjoys the following properties*

(P1) *If X_n converges, its limit is a Nash equilibrium* [M & Zhou, 2019]

(P2) *If x^* is stochastically stable, it is a Nash equilibrium* [Giannou et al., 2021]

(P3) *x^* is stochastically asymptotically stable if and only if it is a strict Nash equilibrium* [Giannou et al., 2021]

(P4) *If $p > 1/2$ and $\mathcal{G}$ is a congestion game, then X_n converges to a Nash equilibrium (a.s.)* [Cohen et al., 2017]

Rate of convergence

Theorem (Giannou et al., 2021)

➥ **Assume:** *all players run (FTRL) with step-size γ_n and oracle parameters b_n (bias) and U_n (noise) as before*

✍ **Then:** *if x^* is a strict Nash equilibrium and X_n converges to x^*, we have*

$$\|X_n - x^*\|_1 \leq \sum_{a \notin \operatorname{supp}(x^*)} \phi\left(A - B \sum_{k=1}^{n} \gamma_k\right)$$

where

- *$A, B > 0$ are initialization- and game-dependent constants*
- *The **rate function** ϕ is determined by the method's regularizer*
  - ***For exponential weights:*** $\phi(z) = \exp(z) \implies$ ***geometric convergence*** *in* $S_n = \sum_{k=1}^{n} \gamma_k$
  - ***For projection dynamics:*** $\phi(z) = [z]_+ \implies$ ***convergence in a finite number of iterations!***

Overview

I. Learning in continuous time

- ✓ Nash equilibrium $\implies$ stationarity
- ✓ Lyapunov stability $\implies$ equilibrium
- ✓ Asymptotic stability $\iff$ strict equilibrium
- ✓ Potential games $\implies$ convergence to equilibrium
- ✓ Zero-sum games $\implies$ Poincaré recurrence

II. Learning in discrete time

- ✗ Depends on feedback, step-size, ...
- ✗ Nash equilibrium $\not\Longrightarrow$ stationarity
- ✓ Lyapunov stability $\implies$ equilibrium
- ✓ Asymptotic stability $\iff$ strict equilibrium
- ✓ Potential games $\implies$ convergence to equilibrium
- ✗ Zero-sum games $\not\Longrightarrow$ Poincaré recurrence

Open questions

- Robustness to delays / corruptions / ...
- Non-singleton attractors? Other limit behaviors?
- Learning in continuous games?

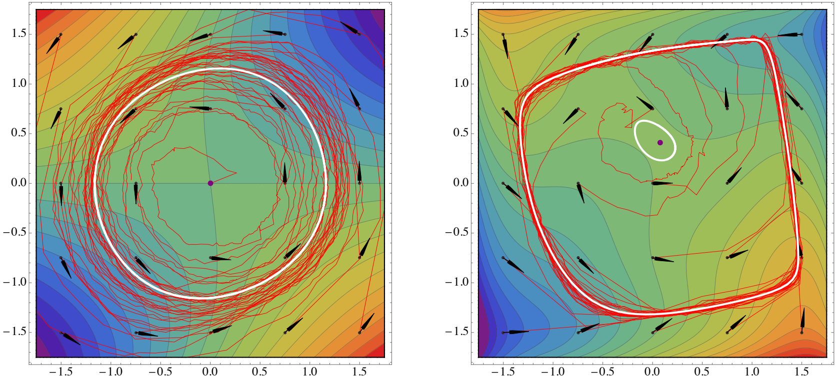


Figure: Limit cycles in almost bilinear games of the form $\min_{x_1 \in \mathcal{X}_1} \max_{x_2 \in \mathcal{X}_2} f(x_1, x_2) = x_1 x_2 + \varepsilon[\phi(x_1) - \phi(x_2)]$

References I

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 1995.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002a.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1): 48–77, 2002b.

Cohen, J., Héliou, A., and Mertikopoulos, P. Learning with bandit feedback in potential games. In *NIPS '17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

Coucheney, P., Gaujal, B., and Mertikopoulos, P. Penalty-regulated dynamics and robust learning procedures in games. *Mathematics of Operations Research*, 40(3):611–633, August 2015.

Flokas, L., Vlatakis-Gkaragkounis, E. V., Lianeas, T., Mertikopoulos, P., and Piliouras, G. No-regret learning and mixed Nash equilibria: They do not mix. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

Freund, Y. and Schapire, R. E. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

Giannou, A., Vlatakis-Gkaragkounis, E. V., and Mertikopoulos, P. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In *COLT '21: Proceedings of the 34th Annual Conference on Learning Theory*, 2021a.

Giannou, A., Vlatakis-Gkaragkounis, E. V., and Mertikopoulos, P. The convergence rate of regularized learning in games: From bandits and uncertainty to optimism and beyond. In *NeurIPS '21: Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021b.

References II

Helbing, D. A mathematical model for behavioral changes by pair interactions. In Haag, G., Mueller, U., and Troitzsch, K. G. (eds.), *Economic Evolution and Demographic Change: Formal Models in Social Sciences*, pp. 330-348. Springer, Berlin, 1992.

Hofbauer, J. and Sigmund, K. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.

Hofbauer, J. and Sigmund, K. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479-519, July 2003.

Hofbauer, J., Sorin, S., and Viossat, Y. Time average replicator and best reply dynamics. *Mathematics of Operations Research*, 34(2):263-269, May 2009.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, Cambridge, UK, 2020.

Littlestone, N. and Warmuth, M. K. The weighted majority algorithm. *Information and Computation*, 108(2):212-261, 1994.

Mertikopoulos, P. and Sandholm, W. H. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4): 1297-1324, November 2016.

Mertikopoulos, P. and Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173 (1-2):465-507, January 2019.

Mertikopoulos, P., Papadimitriou, C. H., and Piliouras, G. Cycles in adversarial regularized learning. In *SODA '18: Proceedings of the 29th annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.

Nash, J. F. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the USA*, 36:48-49, 1950.

Nash, J. F. Non-cooperative games. *The Annals of Mathematics*, 54(2):286-295, September 1951.

References III

Rustichini, A. Optimal properties of stimulus-response learning models. *Games and Economic Behavior*, 29(1-2):244-273, 1999.

Sandholm, W. H. Potential games with continuous player sets. *Journal of Economic Theory*, 97:81-108, 2001.

Schlag, K. H. Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, 78(1):130-156, 1998.

Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107-194, 2011.

Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *NIPS' 06: Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 1265-1272. MIT Press, 2006.

Stampacchia, G. Formes bilineaires coercitives sur les ensembles convexes. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 1964.

Taylor, P. D. and Jonker, L. B. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1-2):145-156, 1978.

Weibull, J. W. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.