ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΠΑΙΓΝΙΩΝ ΚΑΙ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΣΤΑΤΙΣΤΙΚΗΣ & ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

Παναγιώτης Μερτικόπουλος

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Τμήμα Μαθηματικών

Χειμερινό Εξάμηνο, 2023–2024

Outline

Welcome!

Welcome to SEP19: *Topics in Game Theory*

"The study of rational decision-making"

- **Instructors:** Panayotis Mertikopoulos
- **Meeting times:** Mondays 09:00-13:00
- **e-class:** https://eclass.uoa.gr/courses/MATH806/
- **Sessions:** Focus on general theory with some deep dives / practical sessions (TBD)
- **Grading scheme:** split between end-of-term project (50%) and final (50%)

Course overview

Rough breakdown of the course:

1. **Part 1: Basic elements of game theory**
    - Basic notions: Nash equilibrium, dominated strategies,...
    - Basic notions: Nash equilibrium, dominated strategies,...
    - Game classes: potential games, congestion games, price of anarchy,...
    - Game dynamics: replicator dynamics, exponential weights,...

2. **Part 2: Multi-armed bandits and online optimization**
    - Bandits and regret: regret minimization,...
    - Algorithms: Hedge, EXP3,...
    - Online convex optimization: regret, convexification,...
    - Algorithms: leader-following policies, gradient / mirror descent,...

Why game theory?

Example 1: A game of roads

A beautiful morning commute in Chicago

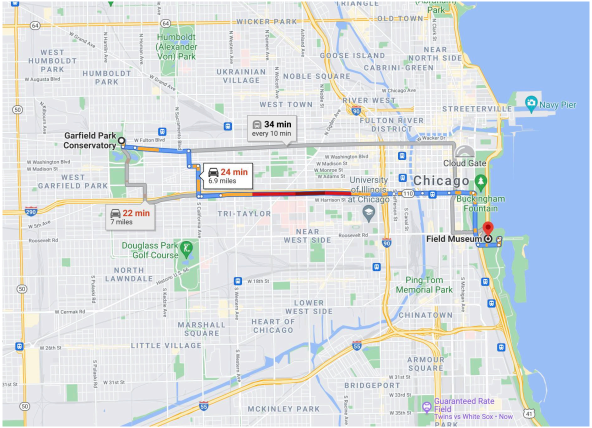
The price of congestion

In the US alone, congestion cost $305 billion in 2017 (≈1.6% of GDP)

source: INRIX

- Lost productivity
- Fuel waste
- Environmental impact, quality of life,...

Game of roads



The city of Chicago

- 2,700,000 people
- 1,261,000 daily trips
- 933 nodes
- 2950 edges
- 870,000 o/d pairs
- $\approx 2 * 10^{16}$ paths

A very large game!

Example 2: Spot the fake

Which person is real?

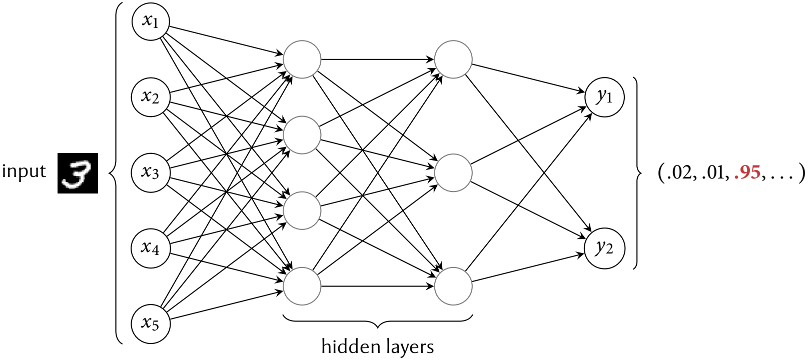
Example 2: Spot the fake

Which person is real?

Spoiler: https://thispersondoesnotexist.com

Neural networks

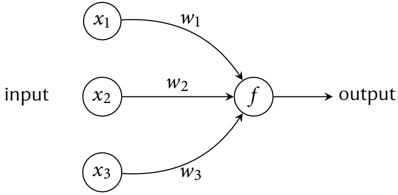
The workhorse of deep learning:



The deep learning revolution: breaking the human perception barrier (2010's)

Neurons

The atoms of any deep learning architecture are its **neurons:**

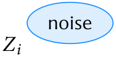


- **Input** could be binary $\{0,1\}$ or real (e.g., average intensity of image)
- Inputs weighed with **weight coefficients** w_i
- Neuron **activates** on value of $f(\sum_i w_i x_i)$

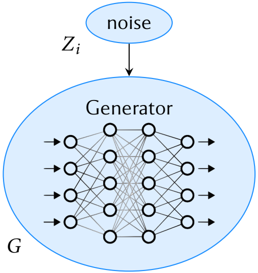
Examples

1. ***Perceptron:*** binary inputs, step function activation
2. ***Sigmoid neuron:*** real inputs, tanh activation
3. ***ReLU:*** real inputs, rectified linear activation ($f(z) = [z]_+$)

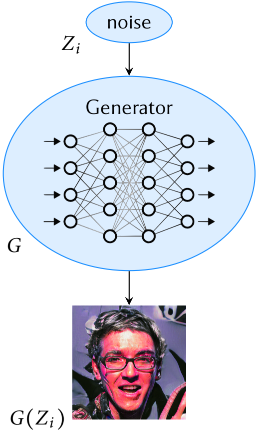
The schematics of GANs



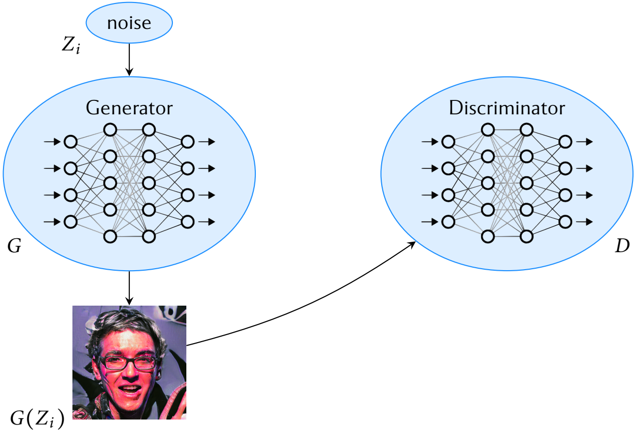
The schematics of GANs



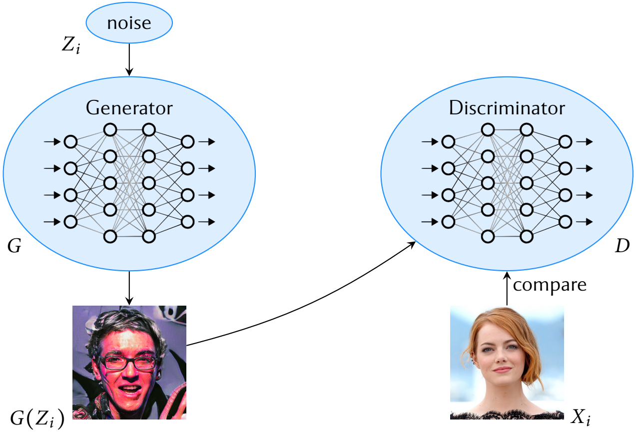
The schematics of GANs



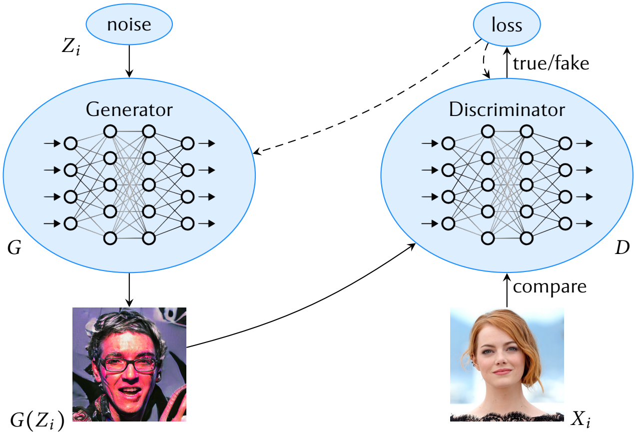
The schematics of GANs



The schematics of GANs



The schematics of GANs



Model likelihood: $\ell(G,D) = \prod_{i=1}^{N} D(X_i) \times \prod_{i=1}^{N}(1 - D(G(Z_i)))$

GAN training

How to find good generators (G) and discriminators (D)?

Discriminator: maximize (log-)likelihood estimation

$$\max_{D \in \mathcal{D}} \log \ell(G, D)$$

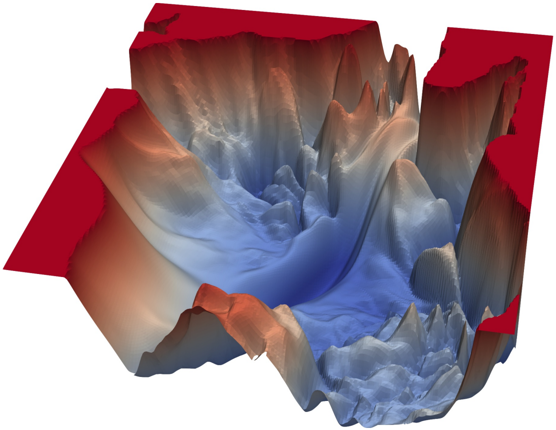
Generator: minimize the resulting divergence

$$\min_{G \in \mathcal{G}} \max_{D \in \mathcal{D}} \log \ell(G, D)$$

A (very complex) zero-sum game!

Training landscape

A deep learning loss landscape



➻ Easier problem: find a needle in a haystack

FailGAN

The game does not always work out:

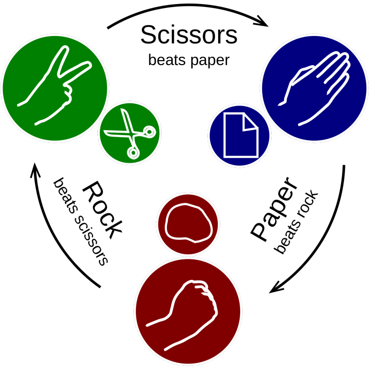
A StyleGAN after 8 days of training at Nvidia headquarters (!!!)

Questions we'll try to answer

1. **How should we model player interactions?**
   - Urban traffic ≠ transit systems ≠ packet networks ≠ …
   - Rational agents ≠ humans ≠ AI algorithms ≠ …
   - Competition ≠ congestion ≠ coordination ≠ …
2. **What is a desired operational state?**
   - Social optimum ≠ equilibrium ≠ …
   - Static (equilibrium, social optimum) ≠ Bayesian ≠ online (regret) ≠ …
3. **How to compute it?**
   - Calculation ≠ learning ≠ implementation
   - Informational constraints: feedback, bounded rationality, uncertainty, …

Outline

Let's play a game



What would you play? How can we model this game mathematically?

Let's play a game, formally

- ***Players:*** "1" and "2"
- ***Actions*** associated to each player: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}, i = 1, 2$
- ***Payoff matrix*** (win: \$1; lose −\$1; tie \$0):

$$A = \begin{array}{c|ccc} & \mathrm{R} & \mathrm{P} & \mathrm{S} \\ \hline \mathrm{R} & 0 & -1 & 1 \\ \mathrm{P} & 1 & 0 & -1 \\ \mathrm{S} & -1 & 1 & 0 \end{array}$$

- ***Payoff functions:***
  - $u_1 \colon \mathcal{A}_1 \times \mathcal{A}_2 \to \mathbb{R}$ given by $u_1(\mathrm{R}, \mathrm{R}) = 0$, $u_1(\mathrm{R}, \mathrm{P}) = -1, \dots$
  - $u_2 \colon \mathcal{A}_1 \times \mathcal{A}_2 \to \mathbb{R}$ given by $u_2(\mathrm{R}, \mathrm{R}) = 0$, $u_2(\mathrm{R}, \mathrm{P}) = 1, \dots$

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*** $u_i \colon \mathcal{A} \equiv \prod_j \mathcal{A}_j \to \mathbb{R}$ 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:

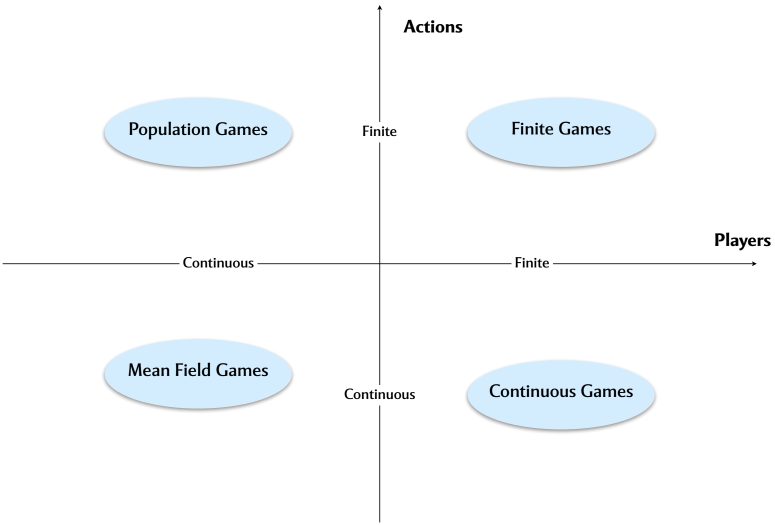
1. A set of ***players*** $\mathcal{N}$
2. A set of ***actions*** (or ***pure strategies***) $\mathcal{A}_i$ per player $i \in \mathcal{N}$
3. An ensemble of ***payoff functions*** $u_i \colon \mathcal{A} \equiv \prod_j \mathcal{A}_j \to \mathbb{R}$ per player $i \in \mathcal{N}$

Important:

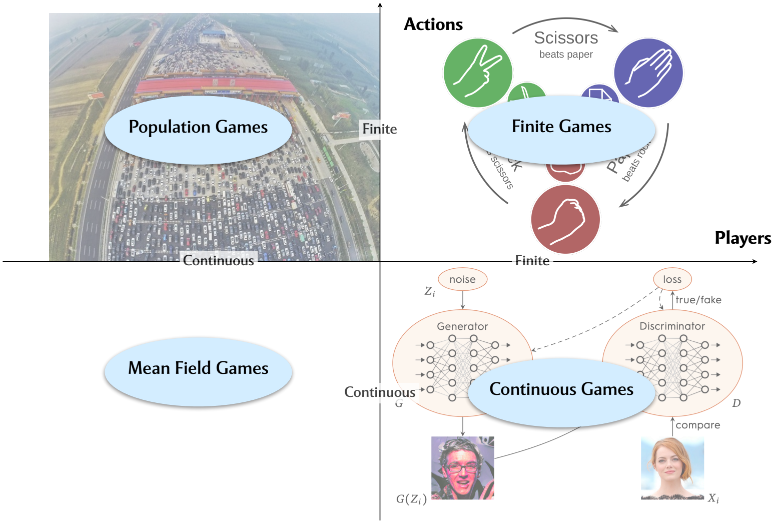
- Player set: atomic vs. nonatomic
- Action sets: finite vs. continuous; shared vs. individual; ...

☞ ***NB:*** do not mix game classes!

Taxonomy


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

Taxonomy


Actions
Players
Finite
Continuous
Finite
Continuous
Population Games
Finite Games
Mean Field Games
Continuous Games
Scissors
beats paper
noise
loss
true/fake
Generator
Discriminator
compare
Z_i
$G(Z_i)$
X_i
D
G

What's in a game?

Definition (Finite games)

A ***finite game in normal form*** is a collection of the following primitives:

- A finite set of ***players*** $\mathcal{N} = \{1, \dots, N\}$
- A finite set of ***actions*** (or ***pure strategies***) $\mathcal{A}_i$ for each player $i \in \mathcal{N}$
- A ***payoff function*** $u_i \colon \mathcal{A} := \prod_j \mathcal{A}_j \to \mathbb{R}$ for each player $i \in \mathcal{N}$

A game with primitives as above will be denoted as $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$.

Some notes:

- "Normal form" $\rightsquigarrow$ difference with "extensive form" games (Chess, Go,...)
- Handy shorthands: $(a_1, \dots, a_i, \dots a_N) \leftarrow (a_i; a_{-i})$ and $\mathcal{A}_{-i} = \prod_{j \neq i} \mathcal{A}_j$

The Prisoner's Dilemma

Bonnie and Clyde are captured by the authorities and put in separate cells:

- If both betray each other, they both serve 2 years in prison
- If Bonnie betrays but Clyde remains silent, Bonnie goes free and Clyde serves 3 years
- If Bonnie remains silent but Clyde betrays, Bonnie serves 3 years and Clyde goes free
- If neither betrays the other, they both serve 1 year

The Prisoner's Dilemma

Bonnie and Clyde are captured by the authorities and put in separate cells:

- If both betray each other, they both serve 2 years in prison
- If Bonnie betrays but Clyde remains silent, Bonnie goes free and Clyde serves 3 years
- If Bonnie remains silent but Clyde betrays, Bonnie serves 3 years and Clyde goes free
- If neither betrays the other, they both serve 1 year

Normal form representation:

- Players: $\mathcal{N} = \{B, C\}$
- Actions: $\mathcal{A}_B = \mathcal{A}_C = \{\texttt{betray}, \texttt{silent}\}$
- Payoff bimatrix:

$B \downarrow \; C \rightarrow$	betray	silent
betray	$(-2, -2)$	$(0, -3)$
silent	$(-3, 0)$	$(-1, -1)$

Split or steal?

`https://www.youtube.com/watch?v=S0qjK3TWZE8`

- If both players steal, they both get nothing
- If one player steals and the other splits, the one who steals gets everything
- If both players split, they split the prize

Do you `split` or `steal`?

Split or steal?

https://www.youtube.com/watch?v=S0qjK3TWZE8

- If both players steal, they both get nothing
- If one player steals and the other splits, the one who steals gets everything
- If both players split, they split the prize

Do you `split` or `steal`?

Normal form representation:

- Players: $\mathcal{N} = \{A, B\}$
- Actions: $\mathcal{A}_A = \mathcal{A}_B = \{\texttt{split}, \texttt{steal}\}$
- Payoff bimatrix:

$A \downarrow \; B \rightarrow$	split	steal
split	$(\$6800, \$6800)$	$(0, \$13600)$
steal	$(\$13600, 0)$	$(0, 0)$

The battle of the sexes

Robin and Charlie want to go out for the evening:

- Robin prefers to go to a movie
- Charlie prefers to go to the theater
- They both prefer being together instead of alone

The battle of the sexes

Robin and Charlie want to go out for the evening:

- Robin prefers to go to a movie
- Charlie prefers to go to the theater
- They both prefer being together instead of alone

Normal form representation:

- Players: $\mathcal{N} = \{R, C\}$
- Actions: $\mathcal{A}_R = \mathcal{A}_C = \{\texttt{movie}, \texttt{theater}\}$
- Payoff bimatrix:

$R \downarrow \; C \rightarrow$	movie	theater
movie	$(3, 2)$	$(0, 0)$
theater	$(0, 0)$	$(2, 3)$

The collision game

Robin and Charlie arrive at an uncontrolled intersection:

- If they both drive through, they crash
- If they both yield, they may wait forever
- If one yields and the other drives through, the latter loses less time

The collision game

Robin and Charlie arrive at an uncontrolled intersection:

- If they both drive through, they crash
- If they both yield, they may wait forever
- If one yields and the other drives through, the latter loses less time

Normal form representation:

- Players: $\mathcal{N} = \{R, C\}$
- Actions: $\mathcal{A}_R = \mathcal{A}_C = \{\texttt{drive}, \texttt{yield}\}$
- Payoff bimatrix:

$R \downarrow\ C \rightarrow$	drive	yield
drive	$(-100, -100)$	$(2, 1)$
yield	$(1, 2)$	$(0, 0)$

Dominated strategies

Sometimes, an action may yield consistently suboptimal payoffs

Definition (Dominated strategies)

1. A strategy $a_i \in \mathcal{A}_i$ is ***strictly dominated*** by $a_i' \in \mathcal{A}_i$ if

$$u_i(a_i; a_{-i}) < u_i(a_i'; a_{-i}) \quad \text{for all } a_{-i} \in \mathcal{A}_{-i}$$

2. A strategy $a_i \in \mathcal{A}_i$ is ***weakly dominated*** by $a_i' \in \mathcal{A}_i$ if

$$u_i(a_i; a_{-i}) \leq u_i(a_i'; a_{-i}) \quad \text{for all } a_{-i} \in \mathcal{A}_{-i}$$

   and $u_i(a_i; a_{-i}) < u_i(a_i'; a_{-i})$ for some $a_{-i} \in \mathcal{A}_{-i}$.

Notation:

- a_i is *strictly* dominated by a_i': $a_i \prec a_i'$
- a_i is *weakly* dominated by a_i': $a_i \preccurlyeq a_i'$

Examples, revisited

The prisoner's dilemma:

$R \downarrow \ C \rightarrow$	betray	silent
betray	$(-2, -2)$	$(0, -3)$
silent	$(-3, 0)$	$(-1, -1)$

Split or steal:

$R \downarrow \ C \rightarrow$	split	steal
split	$(\$6800, \$6800)$	$(0, \$13600)$
steal	$(\$13600, 0)$	$(0, 0)$

Battle of the sexes:

$R \downarrow \ C \rightarrow$	movie	theater
movie	$(3, 2)$	$(0, 0)$
theater	$(0, 0)$	$(2, 3)$

Iteratively dominated strategies

A larger game:

$$\begin{matrix} (9,4) & (5,3) & (3,2) \\ (0,1) & (4,6) & (6,0) \\ (2,1) & (3,5) & (2,4) \end{matrix}$$

Iteratively dominated strategies

A larger game:

$$\begin{matrix} (9,4) & (5,3) & (3,2) \\ (0,1) & (4,6) & (6,0) \\ (2,1) & (3,5) & (2,4) \end{matrix}$$

Definition

1. A strategy is called ***iteratively dominated*** if it becomes dominated after successive elimination of dominated strategies.
2. A game is called ***dominance-solvable*** if the successive elimination of dominated strategies leads to a singleton.

Best responses

What if only the strategy of the opposing player(s) is known?

Definition (Best responses)

The strategy $a_i^* \in \mathcal{A}_i$ is a ***best response*** to $a_{-i} \in \mathcal{A}_{-i}$ if

$$u_i(a_i^*; a_{-i}) \geq u_i(a_i; a_{-i}) \quad \text{for all } a_i \in \mathcal{A}_i$$

or, equivalently, if

$$a_i^* \in \arg\max_{a_i \in \mathcal{A}_i} u_i(a_i; a_{-i}).$$

The set-valued function $\mathrm{BR}_i : \mathcal{A}_{-i} \rightrightarrows \mathcal{A}_i$ given by

$$\mathrm{BR}_i(a_{-i}) = \arg\max_{a_i \in \mathcal{A}_i} u_i(a_i; a_{-i})$$

is called the ***best-response correspondence***.

Examples

The prisoner's dilemma:

$R \downarrow \; C \rightarrow$	betray	silent
betray	$(-2,-2)$	$(0,-3)$
silent	$(-3,0)$	$(-1,-1)$

Split or steal:

$R \downarrow \; C \rightarrow$	split	steal
split	$(\$6800,\$6800)$	$(0,\$13600)$
steal	$(\$13600,0)$	$(0,0)$

Battle of the sexes:

$R \downarrow \; C \rightarrow$	movie	theater
movie	$(3,2)$	$(0,0)$
theater	$(0,0)$	$(2,3)$

Dominated strategies and best responses

Some more examples of best responses

$$\begin{matrix} (9,4) & (5,3) & (3,2) \\ (0,1) & (4,6) & (6,0) \\ (2,1) & (3,5) & (2,8) \end{matrix}$$

Dominated strategies and best responses

Some more examples of best responses

$(9,4)$	$(5,3)$	$(3,2)$
$(0,1)$	$(4,6)$	$(6,0)$
$(2,1)$	$(3,5)$	$(2,8)$

Best responses cannot contain dominated strategies

Dominated strategies and best responses

Some more examples of best responses

$$\begin{matrix} (9,4) & (5,3) & (3,2) \\ (0,1) & (4,6) & (6,0) \\ (2,1) & (3,5) & (2,8) \end{matrix}$$

Best responses cannot contain dominated strategies

➻ What about *weakly* dominated strategies?

Nash equilibrium

Equilibrium: best-responding to each other's actions

Definition (Nash equilibrium)

An action profile $a^* = (a_1^*, \dots, a_N^*)$ is a ***Nash equilibrium*** if

$$a_i^* \in \mathrm{BR}_i(a_{-i}^*) \quad \text{for all } i \in \mathcal{N}$$

or, equivalently, if

$$u_i(a_i^*; a_{-i}^*) \geq u_i(a_i; a_{-i}^*) \quad \text{for all } a_i \in \mathcal{A}_i \text{ and all } i \in \mathcal{N}.$$

Intuition:

- ***Stability:*** no player has an incentive to deviate
- ***Unilateral resilience:*** stable against *individual* player deviations, not multi-player ones

Examples, revisited

The prisoner's dilemma:

$R \downarrow \; C \rightarrow$	betray	silent
betray	$(-2,-2)$	$(0,-3)$
silent	$(-3,0)$	$(-1,-1)$

Split or steal:

$R \downarrow \; C \rightarrow$	split	steal
split	$(\$6800, \$6800)$	$(0, \$13600)$
steal	$(\$13600, 0)$	$(0,0)$

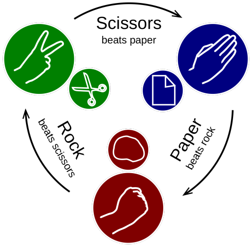
Battle of the sexes:

$R \downarrow \; C \rightarrow$	movie	theater
movie	$(3,2)$	$(0,0)$
theater	$(0,0)$	$(2,3)$

RPS, revisited

How about Rock-Paper-Scissors?

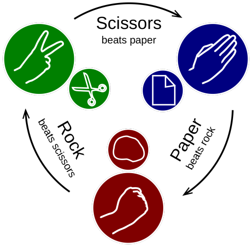
	R	P	S
R	0	−1	1
P	1	0	−1
S	−1	1	0



RPS, revisited

How about Rock-Paper-Scissors?

	R	P	S
R	0	−1	1
P	1	0	−1
S	−1	1	0



Nash equilibria don't always exist!

Mixed strategies

Instead of playing pure strategies, players could ***mix*** their actions:

- ***Mixed strategy*** of player $i \in \mathcal{N}$: probability distribution x_i on $\mathcal{A}_i$
- ***Notation:*** x_{ia_i} = prob. that player i selects $a_i \in \mathcal{A}_i$
- ***Strategy space*** of player i:

$$\mathcal{X}_i := \Delta(\mathcal{A}_i) = \left\{x_i \in \mathbb{R}^{\mathcal{A}_i} : x_{ia_i} \geq 0 \text{ and } \sum_{a_i \in \mathcal{A}_i} x_{ia_i} = 1\right\}$$

➧ $\Delta(\mathcal{A}_i) \rightsquigarrow$ simplex spanned by $\mathcal{A}_i$

- ***Support*** of x_i: actions that are played with positive probability under x_i

$$\operatorname{supp}(x_i) := \{a_i \in \mathcal{A}_i : x_{ia_i} > 0\}$$

- x_i is ***pure*** when $\operatorname{supp}(x_i)$ is a singleton, i.e.,

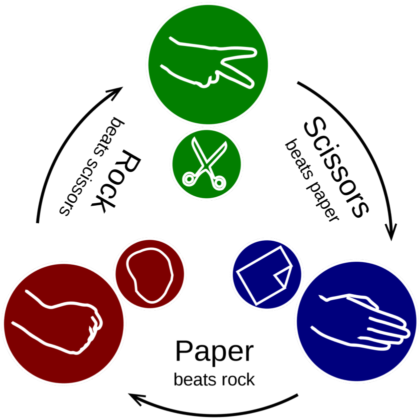
$$\operatorname{supp}(x_i) = \{a_i\} \quad \text{for some } a_i \in \mathcal{A}_i$$

➧ Origin of the term “pure strategies”

RPS, revisited

Playing with mixed strategies:

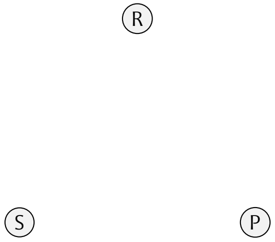
- Players: $\mathcal{N} = \{1, 2\}$



RPS, revisited

Playing with mixed strategies:

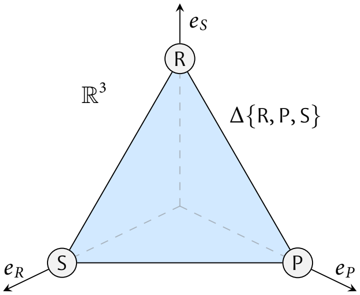
- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{R, P, S\}$



RPS, revisited

Playing with mixed strategies:

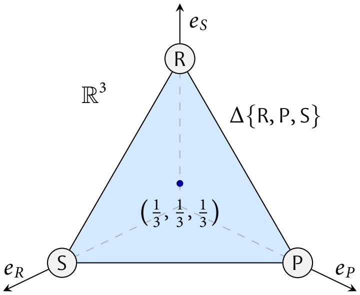
- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{R, P, S\}$
- Mixed strategy space: $\mathcal{X}_i = \Delta\{R, P, S\}$



RPS, revisited

Playing with mixed strategies:

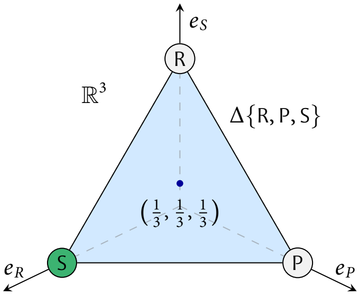
- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Mixed strategy space: $\mathcal{X}_i = \Delta\{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Choose mixed strategy $x_i \in \mathcal{X}_i$



RPS, revisited

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 strategies (collective)

When all players mix their actions:

- Each player $i \in \mathcal{N}$ uses a mixed strategy $x_i \in \mathcal{X}_i$
- Prob. of selecting the action profile $a = (a_1, \dots, a_N) \in \mathcal{A} = \prod_j \mathcal{A}_j$:

$$x_{a_1,\dots,a_N} = \prod_{j \in \mathcal{N}} x_{ja_j}$$

- Prob. of selecting $a_{-i} \in \mathcal{A}_{-i}$:

$$x_{-i;a_{-i}} = \prod_{j \neq i} x_{ja_j}$$

Mixed strategies (collective)

When all players mix their actions:

- Each player $i \in \mathcal{N}$ uses a mixed strategy $x_i \in \mathcal{X}_i$
- Prob. of selecting the action profile $a = (a_1, \dots, a_N) \in \mathcal{A} = \prod_j \mathcal{A}_j$:

$$x_{a_1,\dots,a_N} = \prod_{j \in \mathcal{N}} x_{ja_j}$$

- Prob. of selecting $a_{-i} \in \mathcal{A}_{-i}$:

$$x_{-i;a_{-i}} = \prod_{j \neq i} x_{ja_j}$$

- ***Mixed strategy profile:***

$$x = (x_1, \dots, x_N) \in \mathcal{X} := \prod_{i \in \mathcal{N}} \mathcal{X}_i$$

- ***Mixed strategy profile of i's opponents:***

$$x_{-i} = (x_1, \dots, \cancel{x_i}, \dots, x_N) \in \mathcal{X}_{-i} := \prod_{j \neq i} \mathcal{X}_j$$

☞ **NB:** $\mathcal{X} = \prod_j \Delta(\mathcal{A}_j) \neq \Delta(\prod_j \mathcal{A}_j) = \Delta(\mathcal{A})$

➻ *mixed* vs. *correlated* strategies

Expected payoffs

Expected payoffs under mixed strategies:

- ***Expected payoff to a player*** under a mixed strategy profile:

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

or, in terms of other players' strategies:

$$u_i(x_i; x_{-i}) = \sum_{a_i \in \mathcal{A}_i} \sum_{a_{-i} \in \mathcal{A}_{-i}} x_{i a_i} x_{-i; a_{-i}} \, u_i(a_i; a_{-i})$$

- ***Expected payoff to a pure strategy*** under a mixed strategy profile:

$$v_{i a_i}(x) := u_i(a_i; x_{-i}) = \sum_{a_{-i} \in \mathcal{A}_{-i}} x_{-i; a_{-i}} u_i(a_i; a_{-i})$$

Expected payoffs

Expected payoffs under mixed strategies:

- ***Expected payoff to a player*** under a mixed strategy profile:

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

  or, in terms of other players' strategies:

$$u_i(x_i; x_{-i}) = \sum_{a_i \in \mathcal{A}_i} \sum_{a_{-i} \in \mathcal{A}_{-i}} x_{ia_i} x_{-i;a_{-i}} \, u_i(a_i; a_{-i})$$

- ***Expected payoff to a pure strategy*** under a mixed strategy profile:

$$v_{ia_i}(x) := u_i(a_i; x_{-i}) = \sum_{a_{-i} \in \mathcal{A}_{-i}} x_{-i;a_{-i}} u_i(a_i; a_{-i})$$

- ***Mixed payoff vectors:***

$$v_i(x) = (v_{ia_i}(x))_{a_i \in \mathcal{A}_i} = (u_i(a_i; x_{-i}))_{a_i \in \mathcal{A}_i}$$

  so

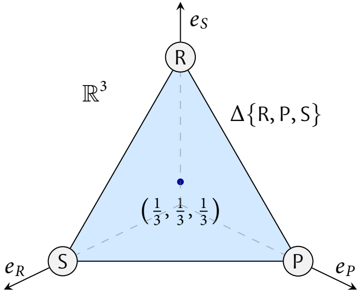
$$u_i(x) = \langle v_i(x), x_i \rangle$$

☞ **NB:** u_i is **linear** in x_i; v_{ia_i} and v_i are **independent** of x_i

Go-to example: Rock-Paper-Scissors

Playing with mixed strategies:

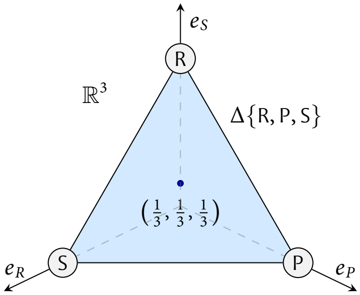
- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{R, P, S\}$
- Mixed strategies: $x_i \in \mathcal{X}_i$



Go-to example: Rock-Paper-Scissors

Playing with mixed strategies:

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



Mixed strategy payoffs:

$$
\begin{aligned}
u_1(x_1, x_2) &= x_{1,\mathrm{R}} x_{2,\mathrm{R}} \cdot (0) + x_{1,\mathrm{R}} x_{2,\mathrm{P}} \cdot (-1) + x_{1,\mathrm{R}} x_{2,\mathrm{S}} \cdot (1) \\
&\quad + x_{1,\mathrm{P}} x_{2,\mathrm{R}} \cdot (1) + x_{1,\mathrm{P}} x_{2,\mathrm{P}} \cdot (0) + x_{1,\mathrm{P}} x_{2,\mathrm{S}} \cdot (-1) \\
&\quad + x_{1,\mathrm{S}} x_{2,\mathrm{R}} \cdot (-1) + x_{1,\mathrm{S}} x_{2,\mathrm{P}} \cdot (1) + x_{1,\mathrm{S}} x_{2,\mathrm{S}} \cdot (0) \\
&= x_{1,\mathrm{R}}(x_{2,\mathrm{S}} - x_{2,\mathrm{P}}) + x_{1,\mathrm{P}}(x_{2,\mathrm{R}} - x_{2,\mathrm{S}}) + x_{1,\mathrm{S}}(x_{2,\mathrm{P}} - x_{2,\mathrm{R}}) \\
&= x_1^\top A x_2 \\
u_2(x_1, x_2) &= -u_1(x_1, x_2)
\end{aligned}
$$

Mixed extensions

Definition (Mixed extension of a finite game)

The ***mixed extension*** of a finite game $\Gamma = \Gamma(\mathcal{N}, \mathcal{A}, u)$ is the ***continuous*** game $\Delta(\Gamma)$ with

- Players $i \in \mathcal{N} = \{1, \dots, N\}$
- Actions $x_i \in \mathcal{X}_i = \Delta(\mathcal{A}_i)$ per player $i \in \mathcal{N}$
- Payoff functions $u_i \colon \mathcal{X} \to \mathbb{R}$, $i \in \mathcal{N}$

Notes:

- ***Continuous game:*** game with *continuous* action spaces (here $\mathcal{X}_i$ instead of $\mathcal{A}_i$)
- ***Context:*** when clear, we will not distinguish between Γ and $\Delta(\Gamma)$

Mixed best responses

Extending the notion of best-responding to mixed strategies

Definition (Mixed best responses)

The mixed strategy $x_i^* \in \mathcal{X}_i$ is a ***best response*** to the mixed profile $x_{-i} \in \mathcal{X}_{-i}$ if

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

or, equivalently, if

$$x_i^* \in \arg\max_{x_i \in \mathcal{X}_i} u_i(x_i; x_{-i}) = \arg\max_{x_i \in \mathcal{X}_i} \langle v_i(x), x_i \rangle$$

As before, we write $\mathrm{BR}_i(x_{-i}) = \arg\max_{x_i \in \mathcal{X}_i} u_i(x_i; x_{-i})$.

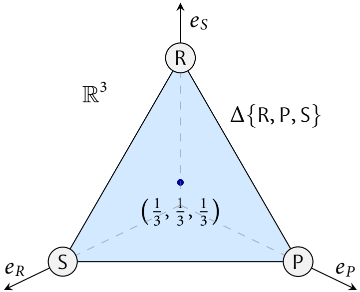
Notes:

- ***Structure:*** $\mathrm{BR}_i(x_{-i})$ is always a face of $\mathcal{X}_i$ ➜ Why?
- ***Notation:*** rely on context to distinguish between pure / mixed best responses

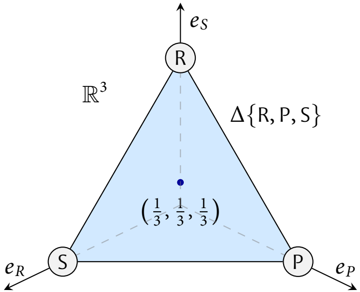
Go-to example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{R, P, S\}$
- Mixed strategies: $x_i^* \in \mathcal{X}_i$



Go-to example: Rock-Paper-Scissors



Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{R, P, S\}$
- Mixed strategies: $x_i^* \in \mathcal{X}_i$

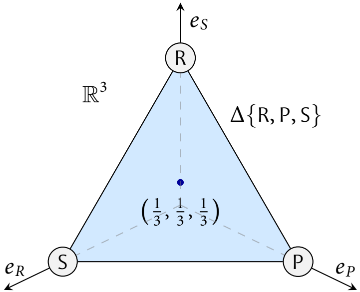
Mixed strategy payoffs when $x_1^* = x_2^* = (1/3, 1/3, 1/3)$:

$$u_1(x_1^*, x_2^*) = \frac{1}{3}\left(\cancel{\frac{1}{3} - \frac{1}{3}}\right) + \frac{1}{3}\left(\cancel{\frac{1}{3} - \frac{1}{3}}\right) + \frac{1}{3}\left(\cancel{\frac{1}{3} - \frac{1}{3}}\right) = 0 = u_2(x_1^*, x_2^*)$$

Go-to example: Rock-Paper-Scissors

Playing with mixed strategies:

- Players: $\mathcal{N} = \{1, 2\}$
- Actions: $\mathcal{A}_i = \{\mathrm{R}, \mathrm{P}, \mathrm{S}\}$
- Mixed strategies: $x_i^* \in \mathcal{X}_i$



Mixed strategy payoffs when $x_1^* = x_2^* = (1/3, 1/3, 1/3)$:

$$u_1(x_1^*, x_2^*) = \tfrac{1}{3}\left(\cancel{\tfrac{1}{3} - \tfrac{1}{3}}\right) + \tfrac{1}{3}\left(\cancel{\tfrac{1}{3} - \tfrac{1}{3}}\right) + \tfrac{1}{3}\left(\cancel{\tfrac{1}{3} - \tfrac{1}{3}}\right) = 0 = u_2(x_1^*, x_2^*)$$

In fact:

$$u_1(x_1, x_2^*) = 0 = u_2(x_1^*, x_2) \quad \text{for all } x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$$

so

$$x_1^* \in \mathrm{BR}_1(x_2^*) \quad \text{and} \quad x_2^* \in \mathrm{BR}_2(x_1^*)$$

Nash equilibrium in mixed strategies

Extending the notion of equilibrium to mixed strategies

Definition (Nash equilibrium)

A strategy profile $x^* = (x_1^*, \dots, x_N^*)$ is a ***Nash equilibrium*** if

$$x_i^* \in \mathrm{BR}_i(x_{-i}^*) \quad \text{for all } i \in \mathcal{N}$$

or, equivalently, if

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

Notes:

- ***Unilateral stability:*** ceteris paribus, no player has an incentive to deviate
- If x^* is pure $\implies$ ***pure Nash equilibrium*** — otherwise "*mixed*"
- If ">" instead of "≥" for $x_i \neq x_i^*$ $\implies$ ***strict Nash equilibrium***

☞ **Prove:** x^* is strict $\iff$ $\mathrm{BR}_i(x_{-i}^*)$ is a singleton for all $i \in \mathcal{N}$

Nash's theorem

RPS admits a Nash equilibrium in mixed strategies – is this always the case?

Nash's theorem

RPS admits a Nash equilibrium in mixed strategies – is this always the case?

Theorem (Nash, 1950)

Every finite game admits a Nash equilibrium in mixed strategies.

Notes:

- ***Support:*** Nash's theorem **does not** specify the support or other properties
- ***Oddness:*** generically odd number of equilibria ➼ Wilson (1971)
- ***Index:*** generically, if m *pure* equilibria, at least $m - 1$ *mixed* equilibria ➼ Ritzberger (1994)

Proof, Part I

Skeleton of the proof:

- Introduce collective best-response correspondence BR: $\mathcal{X} \rightrightarrows \mathcal{X}$ given by

$$\mathrm{BR}(x) = (\mathrm{BR}_i(x_{-i}))_{i=1,\dots,N}$$

- x^* is a Nash equilibrium $\iff$ $x^* \in \mathrm{BR}(x^*)$

Proof, Part I

Skeleton of the proof:

- Introduce collective best-response correspondence $\mathrm{BR}\colon \mathcal{X} \rightrightarrows \mathcal{X}$ given by

$$\mathrm{BR}(x) = (\mathrm{BR}_i(x_{-i}))_{i=1,\dots,N}$$

- x^* is a Nash equilibrium $\iff x^* \in \mathrm{BR}(x^*)$
- Invoke Kakutani's fixed-point theorem for set-valued functions.

Theorem (Kakutani, 1941)

Let $\mathcal{C}$ be a nonempty compact convex subset of $\mathbb{R}^d$, and let $F\colon \mathcal{C} \rightrightarrows \mathcal{C}$ be a set-valued function such that:

(P1) *$F(x)$ is nonempty, closed and convex for all $x \in \mathcal{C}$*

(P2) *F is* **upper hemicontinuous** *at all $x \in \mathcal{C}$, i.e., $\tilde{x} \in F(x)$ whenever $x_t \to x$ and $\tilde{x}_t \to \tilde{x}$ for sequences $x_t \in \mathcal{C}$ and $\tilde{x}_t \in F(x_t)$.*

Then there exists some $x^* \in \mathcal{C}$ such that $x^* \in F(x^*)$.

Upper hemicontinuity ↭ closed graph

Proof, Part II

Verify the conditions of Kakutani's theorem for $\mathcal{C} \leftarrow \mathcal{X}$ and $F \leftarrow \mathrm{BR}$:

(P1) $\mathrm{BR}(x)$ is a face of $\mathcal{X}$, so it is nonempty, closed and convex ➥ Why?

(P2) Argue by contradiction

- Suppose there exist sequences $x_t, \tilde{x}_t \in \mathcal{X}$, $t = 1, 2, \dots$, such that $x_t \to x$, $\tilde{x}_t \to \tilde{x}$ and $\tilde{x}_t \in \mathrm{BR}(x_t)$, but $\tilde{x} \notin \mathrm{BR}(x)$.
- Then there exists a player $i \in \mathcal{N}$ and a deviation $x'_i \in \mathcal{X}_i$ such that

$$u_i(x'_i; x_{-i}) > u_i(\tilde{x}_i; x_{-i})$$

- But since $\tilde{x}_{i,t} \in \mathrm{BR}(x_{-i,t})$ by assumption, we also have:

$$u_i(x'_i; x_{-i,t}) \leq u_i(\tilde{x}_{i,t}; x_{-i,t})$$

- Since $x_t \to x$, $\tilde{x}_t \to \tilde{x}$ and u_i is continuous, taking limits gives

$$u_i(x'_i; x_{-i}) \leq u_i(\tilde{x}_i; x_{-i})$$

which contradicts our original assumption. □

Potential games and best responses

Going back to pure strategies:

- ***In single-player games:*** Nash equilibria (maximizers) trivially exist
- ***In multi-player games:*** not true

Bridge between single- and multi-player settings?

Potential games and best responses

Going back to pure strategies:

- *In single-player games:* Nash equilibria (maximizers) trivially exist
- *In multi-player games:* not true

Bridge between single- and multi-player settings?

Definition (Potential games; Monderer & Shapley, 1996)

A finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ is a *potential game* if there exists a function $\Phi\colon \mathcal{A} \to \mathbb{R}$ such that

$$u_i(a_i'; a_{-i}) - u_i(a_i; a_{-i}) = \Phi(a_i'; a_{-i}) - \Phi(a_i; a_{-i})$$

for all $a, a' \in \mathcal{A}$ and all $i \in \mathcal{N}$.

Potential games and best responses

Going back to pure strategies:

- *In single-player games:* Nash equilibria (maximizers) trivially exist
- *In multi-player games:* not true

Bridge between single- and multi-player settings?

Definition (Potential games; Monderer & Shapley, 1996)

A finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ is a *potential game* if there exists a function $\Phi\colon \mathcal{A} \to \mathbb{R}$ such that

$$u_i(a_i'; a_{-i}) - u_i(a_i; a_{-i}) = \Phi(a_i'; a_{-i}) - \Phi(a_i; a_{-i})$$

for all $a, a' \in \mathcal{A}$ and all $i \in \mathcal{N}$.

Examples

- Battle of the sexes
- Congestion games (more later...)

Basic properties

Existence of equilibria:

- Any ***global maximizer*** $a^* \in \arg\max \Phi$ of Φ is a pure Nash equilibrium

Basic properties

Existence of equilibria:

- Any ***global maximizer*** $a^* \in \arg\max \Phi$ of Φ is a pure Nash equilibrium
- Any ***unilateral maximizer*** $a^* \in \mathcal{A}$ of Φ is a pure Nash equilibrium
- ***Unilateral maximizers:***

$$\Phi(a^*) \geq \Phi(a_i; a^*_{-i}) \quad \text{for all } a_i \in \mathcal{A}_i \text{ and all } i \in \mathcal{N}$$

Basic properties

Existence of equilibria:

- Any ***global maximizer*** $a^* \in \arg\max \Phi$ of Φ is a pure Nash equilibrium
- Any ***unilateral maximizer*** $a^* \in \mathcal{A}$ of Φ is a pure Nash equilibrium
- ***Unilateral maximizers:***

$$\Phi(a^*) \geq \Phi(a_i; a^*_{-i}) \quad \text{for all } a_i \in \mathcal{A}_i \text{ and all } i \in \mathcal{N}$$

When is a game a potential one?

Proposition

Γ is a potential game if and only if

$$\nabla_{x_j} v_i(x) = \nabla_{x_i} v_j(x) \quad \textit{for all } x \in \mathcal{X} \textit{ and all } i, j \in \mathcal{N}$$

where $v_i(x) = (u_i(a_i; x_{-i}))_{a_i \in \mathcal{A}_i}$ is the mixed payoff vector of player $i \in \mathcal{N}$.

Best-response dynamics

A natural updating process:

- Players may choose a new action at each $t = 1, 2, \ldots$
- Players best-respond if this ***strictly*** increases their payoff

Definition (Best-response dynamics)

The ***best-response dynamics*** are defined by the recursion

$$a_{i_t,t+1} \begin{cases} \in \mathrm{BR}_{i_t}(a_{-i_t,t}) & \text{if } a_{i_t,t} \notin \mathrm{BR}_{i_t}(a_{-i_t,t}) \\ = a_{i_t,t} & \text{otherwise} \end{cases} \tag{BRD}$$

where i_t is any player that updates at stage t.

Notes:

- ***Simultaneous:*** all players update simultaneously
- ***Iterative:*** players update in a round robin fashion
- ***Randomized:*** random subset of players updates at any given stage

Convergence

Does (BRD) converge?

Convergence

Does (BRD) converge?

✗ **No - and different modes of updating don't help** ➻ Think RPS

Convergence

Does (BRD) converge?

✗ **No - and different modes of updating don't help** ➧ Think RPS

But good convergence properties in potential games:

Proposition (Monderer & Shapley, 1996)

Let Γ be a finite potential game. Then the iterative version of (BRD) *converges to a pure Nash equilibrium after finitely many steps.*

Convergence

Does (BRD) converge?

✗ **No - and different modes of updating don't help** ➻ Think RPS

But good convergence properties in potential games:

Proposition (Monderer & Shapley, 1996)

Let Γ be a finite potential game. Then the iterative version of (BRD) *converges to a pure Nash equilibrium after finitely many steps.*

Notes:

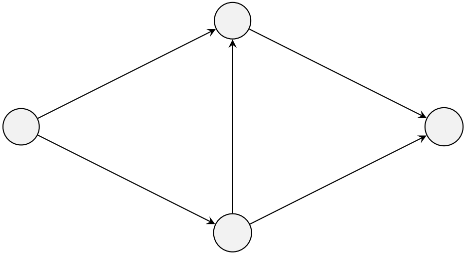
- ***Simple proof:*** potential before and after an update is

$$\Phi(a_i^+; a_{-i}) - \Phi(a_i; a_{-i}) = u_i(a_i^+; a_{-i}) - u_i(a_i; a_{-i}) > 0$$

  whenever $a_i^+ \neq a_i \implies$ no action profile is visited twice $\implies$ the process stops

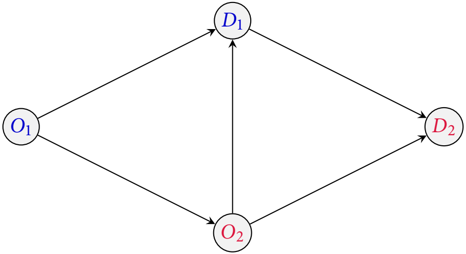
- ***Iterative vs. simultaneous:*** the distinction matters, **simultaneous** (BRD) **may cycle**

Congestion games



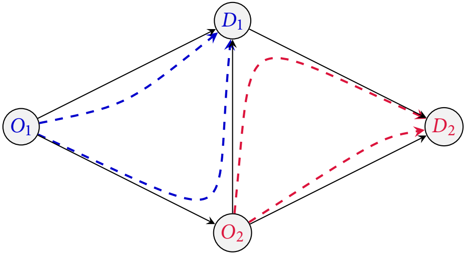
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Congestion games



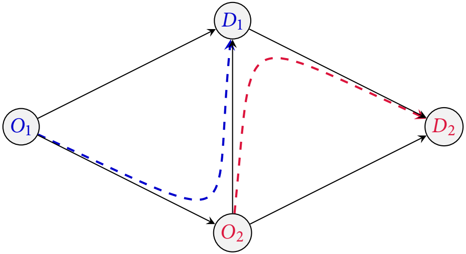
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** i-th player travels from O_i to D_i and induces 1 unit of traffic

Congestion games



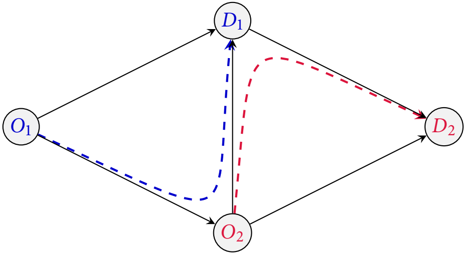
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** i-th player travels from O_i to D_i and induces 1 unit of traffic
- ***Paths $\mathcal{A}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$

Congestion games



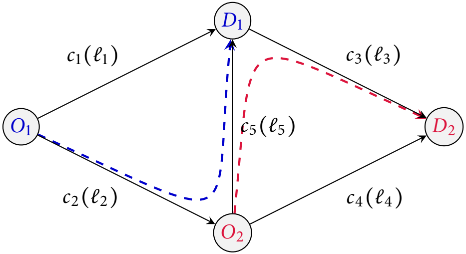
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** i-th player travels from O_i to D_i and induces 1 unit of traffic
- ***Paths $\mathcal{A}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$
- ***Path choice:*** player $i \in \mathcal{N}$ chooses path $a_i \in \mathcal{A}_i$

Congestion games



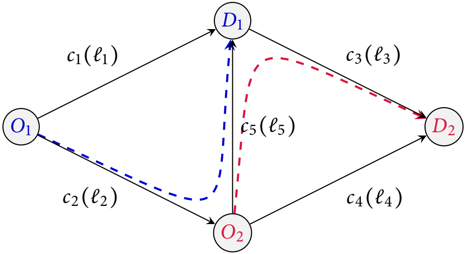
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** i-th player travels from O_i to D_i and induces 1 unit of traffic
- ***Paths $\mathcal{A}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$
- ***Path choice:*** player $i \in \mathcal{N}$ chooses path $a_i \in \mathcal{A}_i$
- ***Load $\ell_e = \sum_{i \in \mathcal{N}} \mathbb{1}(a_i \ni e)$:*** total traffic load along edge e

Congestion games



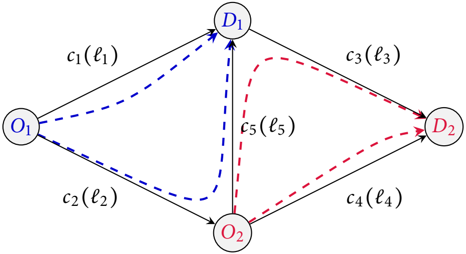
- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs*** $i \in \mathcal{N}$***:*** i-th player travels from O_i to D_i and induces 1 unit of traffic
- ***Paths*** $\mathcal{A}_i$***:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$
- ***Path choice:*** player $i \in \mathcal{N}$ chooses path $a_i \in \mathcal{A}_i$
- ***Load*** $\ell_e = \sum_{i \in \mathcal{N}} \mathbb{1}(a_i \ni e)$***:*** total traffic load along edge e
- ***Edge cost function*** $c_e(\ell_e)$***:*** cost along edge e when edge load is ℓ_e

Congestion games



- ***Network:*** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- ***O/D pairs $i \in \mathcal{N}$:*** i-th player travels from O_i to D_i and induces 1 unit of traffic
- ***Paths $\mathcal{A}_i$:*** (sub)set of paths joining $O_i \rightsquigarrow D_i$
- ***Path choice:*** player $i \in \mathcal{N}$ chooses path $a_i \in \mathcal{A}_i$
- ***Load $\ell_e = \sum_{i \in \mathcal{N}} \mathbb{1}(a_i \ni e)$:*** total traffic load along edge e
- ***Edge cost function $c_e(\ell_e)$:*** cost along edge e when edge load is ℓ_e
- ***Player cost:*** $c_i(a) = \sum_{e \in a_i} c_e(\ell_e)$

Congestion games



- **Network:** multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- **O/D pairs $i \in \mathcal{N}$:** i-th player travels from O_i to D_i and induces 1 unit of traffic
- **Paths $\mathcal{A}_i$:** (sub)set of paths joining $O_i \rightsquigarrow D_i$
- **Path choice:** player $i \in \mathcal{N}$ chooses path $a_i \in \mathcal{A}_i$
- **Load $\ell_e = \sum_{i \in \mathcal{N}} \mathbb{1}(a_i \ni e)$:** total traffic load along edge e
- **Edge cost function $c_e(\ell_e)$:** cost along edge e when edge load is ℓ_e
- **Player cost:** $c_i(a) = \sum_{e \in a_i} c_e(\ell_e)$
- **Congestion game (atomic, non-splittable):** $\Gamma = (\mathcal{G}, \mathcal{N}, \mathcal{A}, c)$

Rosenthal Potential

Potential games

- Potential function: $\Phi(a_i'; a_{-i}) - \Phi(a_i; a_{-i}) = u_i(a_i'; a_{-i}) - u_i(a_i; a_{-i})$ for all $a_i, a_i' \in \mathcal{A}_i$.
- Pure equilibria exist and can be found by best-response dynamics

Rosenthal Potential

Potential games

- Potential function: $\Phi(a_i'; a_{-i}) - \Phi(a_i; a_{-i}) = u_i(a_i'; a_{-i}) - u_i(a_i; a_{-i})$ for all $a_i, a_i' \in \mathcal{A}_i$.
- Pure equilibria exist and can be found by best-response dynamics

Theorem (Rosenthal, 1973)

Any (atomic, non-splittable) congestion game admits the potential function

$$\Phi(a) = \sum_{e \in \mathcal{E}} \sum_{k=1}^{\ell_e(a)} c_e(k) \quad \text{for all } a \in \prod_{i \in \mathcal{N}} \mathcal{A}_i$$

Proof of Rosenthal's Theorem

Theorem (Rosenthal, 1973)

Any (atomic, non-splittable) congestion game admits the potential function

$$\Phi(a) = \sum_{e \in \mathcal{E}} \sum_{k=1}^{\ell_e(a)} c_e(k) \quad \textit{for all } a \in \prod_{i \in \mathcal{N}} \mathcal{A}_i$$

Proof.

Consider a strategy profile $a \in \prod_{i \in \mathcal{N}} \mathcal{A}_i$ and a strategy $a'_i \in \mathcal{A}_i$. Then:

$$\begin{aligned}
\Phi(a'_i; a_{-i}) - \Phi(a_i; a_{-i}) &= \sum_{e \in \mathcal{E}} \sum_{k=1}^{\ell_e(a'_i; a_{-i})} c_e(k) - \sum_{e \in \mathcal{E}} \sum_{k=1}^{\ell_e(a_i, -a_i)} c_e(k) \\
&= \sum_{e \in a'_i \setminus a_i} c_e(\ell_e(a) + 1) - \sum_{e \in a_i \setminus a'_i} c_e(\ell_e(a)). \\
&= \dots
\end{aligned}$$

□

NB: The converse is also true (Monderer & Shapley, 1996).

The Price of Anarchy

How bad is selfish routing?

The Price of Anarchy

How bad is selfish routing?

Definition (Social optimum)

The ***social optimum*** of a congestion game is the value

$$\mathrm{Opt}(\Gamma) = \min_{a \in \mathcal{A}} C(a) \tag{SO}$$

where $C(a) = \sum_{i \in \mathcal{N}} c_i(a)$ is the game's ***social cost*** function.

Definition (Price of Anarchy; Koutsoupias & Papadimitriou, 1999)

The ***POA! (POA!)*** of a congestion game Γ is defined as

$$\mathrm{PoA}(\Gamma) = \max_{a^* \in \mathrm{Eq}(\Gamma)} \frac{C(a^*)}{\mathrm{Opt}(\Gamma)}. \tag{PoA}$$

The Braess network

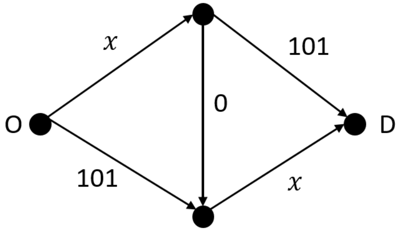


Figure: The Braess network

Bounds of PoA: Linear costs I

We will focus on the games with **linear costs**, i.e., $c_e(\ell) = A_e\ell + B_e, \forall e$.

Theorem (Christodoulou & Koutsoupias '05)

In any (nonatomic splittable) congestion game with linear cost functions $\mathrm{PoA}(\Gamma) \leq \frac{5}{2}$.

- **NB:** focus for simplicity on the ***identity cost*** function $c_e(\ell) = \ell$
  - Let a^* be any equilibrium and a^{Opt} be an action minimizing the social cost:

$$c_i(a_i^*, a_{-i}^*) \leq c_i(a_i^{\mathrm{Opt}}, a_{-i}^*) = \sum_{e \in a_i^{\mathrm{Opt}}} c_e(\ell_e(a_i^{\mathrm{Opt}}, a_{-i}^*)) \leq \sum_{e \in a_i^{\mathrm{Opt}}} c_e(\ell_e(a^*) + 1)$$

  - Then:

$$C(a^*) = \sum_{i \in \mathcal{N}} c_i(a^*) \leq \sum_{i \in \mathcal{N}} \sum_{e \in a_i^{\mathrm{Opt}}} c_e(\ell_e(a^*) + 1) = \sum_{e \in \mathcal{E}} \ell_e(a^{\mathrm{Opt}}) \cdot [\ell_e(a^*) + 1]$$

  - The social cost may further be bounded as

$$C(a^*) \leq \sum_{e \in \mathcal{E}} \frac{[\ell_e(a^{\mathrm{Opt}})]^2}{3} + \frac{5[\ell_e(a^{\mathrm{Opt}})]^2}{3} = \frac{1}{3}C(a^*) + \frac{5}{3}C(a^{\mathrm{Opt}})$$

Bounds of PoA: Linear costs II

- **NB:** For any *positive* integers α, β, we have $\beta(\alpha+1) \leq \frac{\alpha^2}{3} + \frac{5\beta^2}{3}$.
  - Similar analysis for linear cost ($h_e \neq 1, k_e \neq 0$). □

Outline

Basic questions

How do players learn from the history of play?

Do players end up playing a Nash equilibrium?

The model

Sequence of events

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

repeat

At each epoch $t \geq 0$ **do simultaneously** for all players $i \in \mathcal{N}$ # continuous time

Choose ***mixed strategy*** $x_i(t) \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # mixing

Encounter ***mixed payoff vector*** $v_i(x(t))$ and get **mixed payoff** $u_i(x(t)) = \langle v_i(t), x(t) \rangle$ # feedback phase

until end

The model

Sequence of events

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

repeat

At each epoch $t \geq 0$ **do simultaneously** for all players $i \in \mathcal{N}$ — # continuous time

Choose ***mixed strategy*** $x_i(t) \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ — # mixing

Encounter ***mixed payoff vector*** $v_i(x(t))$ and get **mixed payoff** $u_i(x(t)) = \langle v_i(t), x(t) \rangle$ — # feedback phase

until end

Defining elements

- ***Time:*** continuous
- ***Players:*** finite
- ***Actions:*** finite
- ***Mixing:*** yes
- ***Feedback:*** mixed payoff vectors

Exponential weights

Exponential reinforcement mechanism:

- Score each action based on its cumulative payoff over time:

$$y_{ia_i}(t) = \int_0^t v_{ia_i}(x(s))\, ds$$

- Play an action with probability exponentially proportional to its score

$$x_{ia_i}(t) \propto \exp(y_{ia_i}(t))$$

Exponential weight dynamics

$$\begin{aligned} \dot{y}_{ia_i} &= v_{ia_i}(x) \\ x_{ia_i} &= \frac{\exp(y_{ia_i})}{\sum_{a'_i \in \mathcal{A}_i} \exp(y_{ia'_i})} \end{aligned} \tag{EW}$$

The replicator dynamics

How do mixed strategies evolve under (EWD)?

The replicator dynamics

How do mixed strategies evolve under (EWD)?

The replicator dynamics (Taylor & Jonker, 1978)

$$\begin{aligned} \dot{x}_{ia_i} &= x_{ia_i}\left[v_{ia_i}(x) - \sum\nolimits_{a_i' \in \mathcal{A}_i} x_{ia_i'} v_{ia_i'}(x)\right] \\ &= x_{ia_i}\left[u_i(a_i; x_{-i}) - u_i(x)\right] \end{aligned} \tag{RD}$$

"The per capita growth rate of a strategy is proportional to its payoff excess"

➻ Hofbauer & Sigmund (1998); Weibull (1995); Hofbauer & Sigmund (2003); Sandholm (2010)

The replicator dynamics

How do mixed strategies evolve under (EWD)?

The replicator dynamics (Taylor & Jonker, 1978)

$$\begin{aligned} \dot{x}_{ia_i} &= x_{ia_i}\left[v_{ia_i}(x) - \textstyle\sum_{a_i' \in \mathcal{A}_i} x_{ia_i'} v_{ia_i'}(x)\right] \\ &= x_{ia_i}\left[u_i(a_i; x_{-i}) - u_i(x)\right] \end{aligned} \tag{RD}$$

"The per capita growth rate of a strategy is proportional to its payoff excess"

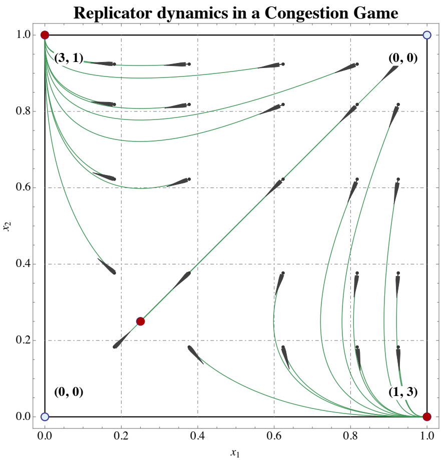
➥ Hofbauer & Sigmund (1998); Weibull (1995); Hofbauer & Sigmund (2003); Sandholm (2010)

Proposition

Solution orbits of (EWD) $\Longleftrightarrow$ *interior orbits of* (RD)

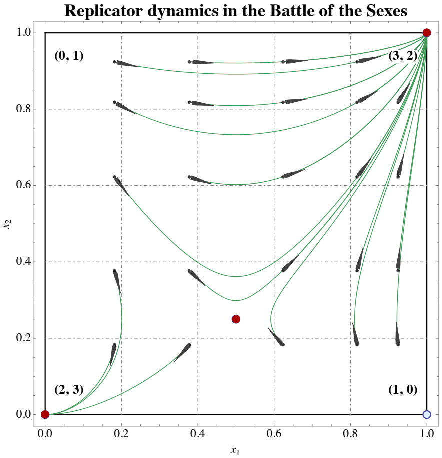
Evolution of mixed strategies: Examples

What do the dynamics look like?



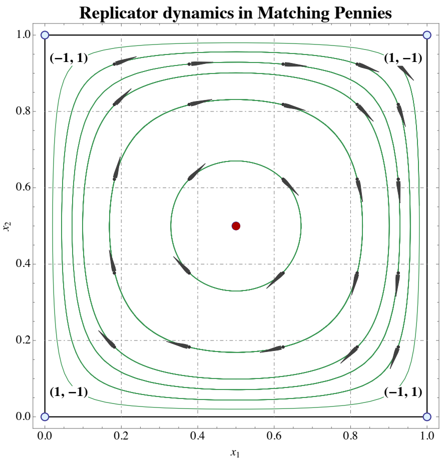
Evolution of mixed strategies: Examples

What do the dynamics look like?



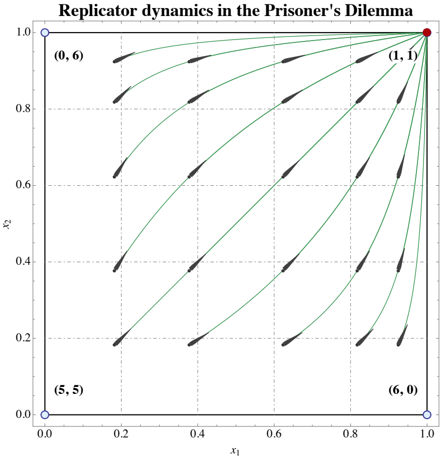
Evolution of mixed strategies: Examples

What do the dynamics look like?



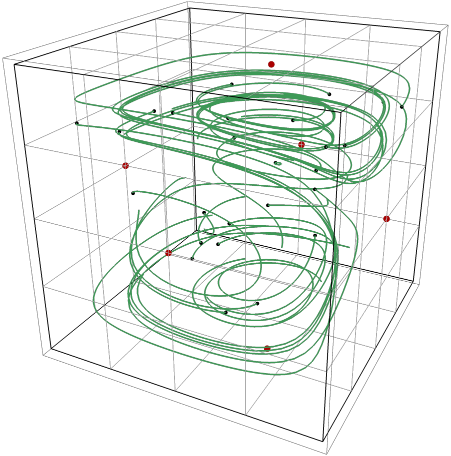
Evolution of mixed strategies: Examples

What do the dynamics look like?



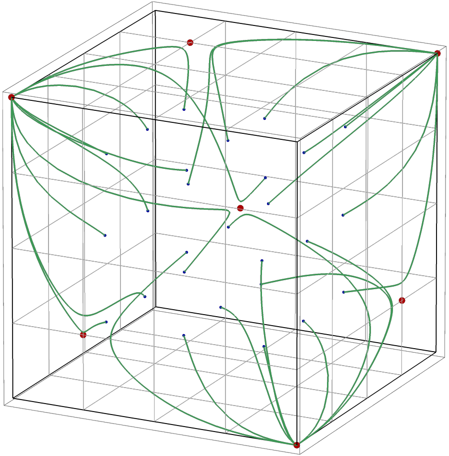
Evolution of mixed strategies: Examples

What do the dynamics look like?



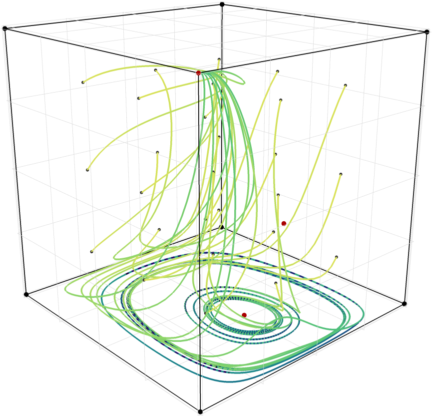
Evolution of mixed strategies: Examples

What do the dynamics look like?



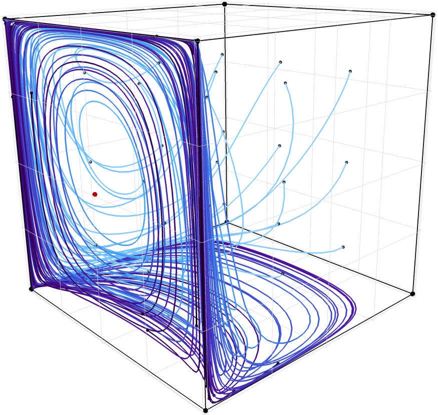
Evolution of mixed strategies: Examples

What do the dynamics look like?



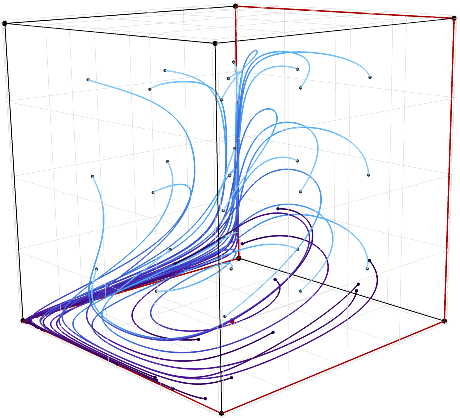
Evolution of mixed strategies: Examples

What do the dynamics look like?



Evolution of mixed strategies: Examples

What do the dynamics look like?



Structural properties

Basic properties of (EWD) / (RD)

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

Proof: Picard-Lindelöf

Structural properties

Basic properties of (EWD) / (RD)

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

  ➥ Proof: Picard-Lindelöf

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

  ➥ Assuming $x(0) \in \mathcal{X}$

Structural properties

Basic properties of (EWD) / (RD)

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

  ➠ Proof: Picard–Lindelöf

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

  ➠ Assuming $x(0) \in \mathcal{X}$

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

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

Dynamics and rationality

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

- Do dominated strategies die out in the long run?
- Are Nash equilibria stationary?
- Are they ***stable?*** Are they ***attracting?***
- Do the replicator dynamics always converge?
- What other behaviors can we observe?
- …

Dominated strategies

Suppose $a_i \in \mathcal{A}_i$ is ***dominated*** by $a_i' \in \mathcal{A}_i$

- Consistent payoff gap:

$$v_{ia_i}(x) \leq v_{ia_i'}(x) - \varepsilon \quad \text{for some } \varepsilon > 0$$

Dominated strategies

Suppose $a_i \in \mathcal{A}_i$ is **dominated** by $a_i' \in \mathcal{A}_i$

- Consistent payoff gap:

$$v_{ia_i}(x) \leq v_{ia_i'}(x) - \varepsilon \quad \text{for some } \varepsilon > 0$$

- Consistent difference in scores:

$$y_{ia_i}(t) = \int_0^t v_{ia_i}(x)\,ds \leq \int_0^t [v_{ia_i'}(x) - \varepsilon]\,ds = y_{ia_i'}(t) - \varepsilon t$$

Dominated strategies

Suppose $a_i \in \mathcal{A}_i$ is ***dominated*** by $a_i' \in \mathcal{A}_i$

- Consistent payoff gap:

$$v_{ia_i}(x) \leq v_{ia_i'}(x) - \varepsilon \quad \text{for some } \varepsilon > 0$$

- Consistent difference in scores:

$$y_{ia_i}(t) = \int_0^t v_{ia_i}(x)\, ds \leq \int_0^t [v_{ia_i'}(x) - \varepsilon]\, ds = y_{ia_i'}(t) - \varepsilon t$$

- Consistent difference in choice probabilities

$$\frac{x_{ia_i}(t)}{x_{ia_i'}(t)} = \frac{\exp(y_{ia_i}(t))}{\exp(y_{ia_i'}(t))} \leq \exp(-\varepsilon t)$$

Dominated strategies

Suppose $a_i \in \mathcal{A}_i$ is **dominated** by $a'_i \in \mathcal{A}_i$

- Consistent payoff gap:

$$v_{ia_i}(x) \leq v_{ia'_i}(x) - \varepsilon \quad \text{for some } \varepsilon > 0$$

- Consistent difference in scores:

$$y_{ia_i}(t) = \int_0^t v_{ia_i}(x)\, ds \leq \int_0^t [v_{ia'_i}(x) - \varepsilon]\, ds = y_{ia'_i}(t) - \varepsilon t$$

- Consistent difference in choice probabilities

$$\frac{x_{ia_i}(t)}{x_{ia'_i}(t)} = \frac{\exp(y_{ia_i}(t))}{\exp(y_{ia'_i}(t))} \leq \exp(-\varepsilon t)$$

Theorem (Samuelson & Zhang (1992))

Let $x(t)$ be a solution orbit of (EWD)/(RD). If $a_i \in \mathcal{A}_i$ is dominated, then

$$x_{ia_i}(t) = \exp(-\Theta(t)) \quad \text{as } t \to \infty$$

In words: under (EWD)/(RD), dominated strategies become extinct at an exponential rate.

Dominated strategies

Suppose $a_i \in \mathcal{A}_i$ is ***dominated*** by $a_i' \in \mathcal{A}_i$

- Consistent payoff gap:

$$v_{ia_i}(x) \leq v_{ia_i'}(x) - \varepsilon \quad \text{for some } \varepsilon > 0$$

- Consistent difference in scores:

$$y_{ia_i}(t) = \int_0^t v_{ia_i}(x)\, ds \leq \int_0^t [v_{ia_i'}(x) - \varepsilon]\, ds = y_{ia_i'}(t) - \varepsilon t$$

- Consistent difference in choice probabilities

$$\frac{x_{ia_i}(t)}{x_{ia_i'}(t)} = \frac{\exp(y_{ia_i}(t))}{\exp(y_{ia_i'}(t))} \leq \exp(-\varepsilon t)$$

Theorem (Samuelson & Zhang (1992))

Let $x(t)$ be a solution orbit of (EWD)/(RD)*. If $a_i \in \mathcal{A}_i$ is dominated, then*

$$x_{ia_i}(t) = \exp(-\Theta(t)) \quad \textit{as } t \to \infty$$

In words: under (EWD)/(RD)*, dominated strategies become extinct at an exponential rate.*

➬ **Self-check:** extend to *iteratively* dominated strategies

Stationarity of equilibria

Nash 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:

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

Stationarity of equilibria

Nash 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:

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

Stationarity of equilibria

Nash 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

Nash 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

Nash 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!**

➬ **Self-check:** All vertices of $\mathcal{X}$ are stationary. General statement?

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 and equilibrium

Proposition (Folk)

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

Stability and equilibrium

Proposition (Folk)

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

Proof. Argue by contradiction:

- **Suppose that x^* is not Nash.** Then

$$v_{ia_i^*}(x^*) = u_i(a_i^*; x_{-i}^*) < u_i(a_i; x_{-i}^*) = v_{ia_i}(x^*)$$

for some $a_i^* \in \operatorname{supp}(x_i^*)$, $a_i \in \mathcal{A}_i$, $i \in \mathcal{N}$

Stability and equilibrium

Proposition (Folk)

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

Proof. Argue by contradiction:

- **Suppose that x^* is not Nash.** Then

$$v_{ia_i^*}(x^*) = u_i(a_i^*; x_{-i}^*) < u_i(a_i; x_{-i}^*) = v_{ia_i}(x^*)$$

  for some $a_i^* \in \operatorname{supp}(x_i^*)$, $a_i \in \mathcal{A}_i$, $i \in \mathcal{N}$

- There exist $\varepsilon > 0$ and neighborhood $\mathcal{U}$ of x^* such that $v_{ia_i}(x) - v_{ia_i^*}(x) > \varepsilon$ for $x \in \mathcal{U}$

Stability and equilibrium

Proposition (Folk)

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

Proof. Argue by contradiction:

- **Suppose that x^* is not Nash.** Then

$$v_{ia_i^*}(x^*) = u_i(a_i^*; x_{-i}^*) < u_i(a_i; x_{-i}^*) = v_{ia_i}(x^*)$$

for some $a_i^* \in \operatorname{supp}(x_i^*)$, $a_i \in \mathcal{A}_i$, $i \in \mathcal{N}$

- There exist $\varepsilon > 0$ and neighborhood $\mathcal{U}$ of x^* such that $v_{ia_i}(x) - v_{ia_i^*}(x) > \varepsilon$ for $x \in \mathcal{U}$

- If $x(t)$ is contained in $\mathcal{U}$ for all $t \geq 0$ **(Lyapunov property)**, then:

$$y_{ia_i^*}(t) - y_{ia_i}(t) = c + \int_0^t [v_{ia_i^*}(x(s)) - v_{ia_i}(x(s))]\, ds < c - \varepsilon t$$

Stability and equilibrium

Proposition (Folk)

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

Proof. Argue by contradiction:

- **Suppose that x^* is not Nash.** Then

$$v_{ia_i^*}(x^*) = u_i(a_i^*; x_{-i}^*) < u_i(a_i; x_{-i}^*) = v_{ia_i}(x^*)$$

  for some $a_i^* \in \operatorname{supp}(x_i^*)$, $a_i \in \mathcal{A}_i$, $i \in \mathcal{N}$

- There exist $\varepsilon > 0$ and neighborhood $\mathcal{U}$ of x^* such that $v_{ia_i}(x) - v_{ia_i^*}(x) > \varepsilon$ for $x \in \mathcal{U}$

- If $x(t)$ is contained in $\mathcal{U}$ for all $t \geq 0$ (**Lyapunov property**), then:

$$y_{ia_i^*}(t) - y_{ia_i}(t) = c + \int_0^t [v_{ia_i^*}(x(s)) - v_{ia_i}(x(s))]\, ds < c - \varepsilon t$$

- We conclude that $x_{ia_i^*}(t) \to 0$, contradicting the Lyapunov stability of x^*. □

Asymptotic stability

Are Nash equilibria attracting?

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 Nash equilibria attracting?

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).

Asymptotic stability

Are Nash equilibria attracting?

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).

Proof. Compare scores:

- If $a^* = (a_1^*, \dots, a_N^*)$ is strict Nash $\implies v_{ia_i^*}(x^*) > v_{ia_i}(x^*)$ for all $a_i \in \mathcal{A}_i \setminus \{a_i^*\}$
- There exist $\varepsilon > 0$ and a nhd $\mathcal{U}$ of x^* such that $v_{ia_i^*}(x) - v_{ia_i}(x) > \varepsilon$ for $x \in \mathcal{U}$

Asymptotic stability

Are Nash equilibria attracting?

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).

Proof. Compare scores:

- If $a^* = (a_1^*, \dots, a_N^*)$ is strict Nash $\implies v_{ia_i^*}(x^*) > v_{ia_i}(x^*)$ for all $a_i \in \mathcal{A}_i \setminus \{a_i^*\}$
- There exist $\varepsilon > 0$ and a nhd $\mathcal{U}$ of x^* such that $v_{ia_i^*}(x) - v_{ia_i}(x) > \varepsilon$ for $x \in \mathcal{U}$
- If $x(t)$ remains in $\mathcal{U}$ for all $t \geq 0$, then

$$y_{ia_i}(t) - y_{ia_i^*}(t) = c + \int_0^t [v_{ia_i}(x(s)) - v_{ia_i^*}(x(s))]\, ds < c - \varepsilon t$$

i.e., $\lim_{t\to\infty} x_{ia_i}(t) = 0$

Asymptotic stability

Are Nash equilibria attracting?

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).

Proof. Compare scores:

- If $a^* = (a_1^*, \dots, a_N^*)$ is strict Nash $\implies v_{ia_i^*}(x^*) > v_{ia_i}(x^*)$ for all $a_i \in \mathcal{A}_i \setminus \{a_i^*\}$
- There exist $\varepsilon > 0$ and a nhd $\mathcal{U}$ of x^* such that $v_{ia_i^*}(x) - v_{ia_i}(x) > \varepsilon$ for $x \in \mathcal{U}$
- If $x(t)$ remains in $\mathcal{U}$ for all $t \geq 0$, then

$$y_{ia_i}(t) - y_{ia_i^*}(t) = c + \int_0^t [v_{ia_i}(x(s)) - v_{ia_i^*}(x(s))]\, ds < c - \varepsilon t$$

  i.e., $\lim_{t\to\infty} x_{ia_i}(t) = 0$

- Proof complete by showing Lyapunov stability ➧ Left as self-check exercise □

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*

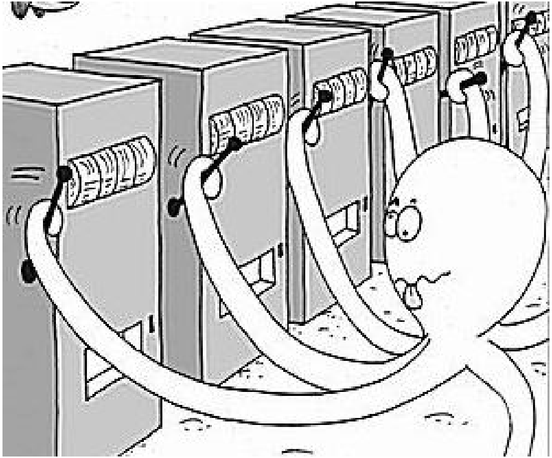
Notes:

- ✗ **Converse to (1), (2) and (3) does not hold!**
- ✓ Proof of (2) similar to (3) — Do as self-check
- Proof of "$\Longleftarrow$" in (4): requires different techniques

Outline

Multi-armed bandits

Robbins' multi-armed bandit problem: **how to play in a (rigged) casino?**



Game-theoretic learning

Sequence of events – continuous time

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

repeat

At each epoch $t \geq 0$ **do simultaneously** for all players $i \in \mathcal{N}$ # continuous time

Choose ***mixed strategy*** $x_i(t) \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # mixing

Encounter ***mixed payoff vector*** $v_i(x(t))$ and get **mixed payoff** $u_i(x(t)) = \langle v_i(t), x(t) \rangle$ # feedback phase

until end

Defining elements

- **Time:** $t \geq 0$
- **Players:** finite
- **Actions:** finite
- **Payoffs:** game
- **Feedback:** mixed payoff vectors

Online learning

Sequence of events – continuous time

Require: set of actions $\mathcal{A} = \{1, \dots, A\}$, stream of payoff vectors $v_t \in [0,1]^{\mathcal{A}}$, $t \geq 0$

repeat

At each epoch $t \geq 0$ **do** # continuous time

Choose ***mixed strategy*** $x_t \in \mathcal{X}$ # mixing

Encounter ***payoff vector*** v_t and get ***mixed payoff*** $u_t(x_t) = \langle v_t, x_t \rangle$ # feedback phase

until end

Defining elements

- **Time:** $t \geq 0$
- **Players:** ***single*** # *"unilateral viewpoint"*
- **Actions:** finite
- **Payoffs:** ***exogenous*** # *"game against Nature"*
- **Feedback:** mixed payoff vectors

Online v. multi-agent learning

How are payoffs generated?

- **Multi-agent viewpoint**
  - *Multiple agents*
  - *Endogenous rewards:* individual payoffs depend on other agents
  - *Game-theoretic:* underlying mechanism is a (finite) game
- **Online viewpoint**
  - *Single agent*
  - *Exogenous rewards:* different payoff vector at each stage
  - *Agnostic:* no assumptions on mechanism generating $v(t)$ # dispassionate Nature

Online v. multi-agent learning

How are payoffs generated?

- **Multi-agent viewpoint**
  - *Multiple agents*
  - *Endogenous rewards:* individual payoffs depend on other agents
  - *Game-theoretic:* underlying mechanism is a (finite) game
- **Online viewpoint**
  - *Single agent*
  - *Exogenous rewards:* different payoff vector at each stage
  - *Agnostic:* no assumptions on mechanism generating $v(t)$ # dispassionate Nature

What is the interplay between online and multi-agent learning?

The agent's regret

Performance of a policy x_t measured by the agent's **regret**

$$u_t(p) - u_t(x_t)$$

The agent's regret

Performance of a policy x_t measured by the agent's **regret**

$$\int_0^T [u_t(p) - u_t(x_t)]\, dt$$

The agent's regret

Performance of a policy x_t measured by the agent's **regret**

$$\max_{p \in \mathcal{X}} \int_0^T [u_t(p) - u_t(x_t)]\, dt$$

The agent's regret

Performance of a policy x_t measured by the agent's **regret**

$$\operatorname{Reg}(T) = \max_{p \in \mathcal{X}} \int_0^T [u_t(p) - u_t(x_t)]\, dt = \max_{p \in \mathcal{X}} \int_0^T \langle v_t, p - x_t \rangle\, dt$$

The agent's regret

Performance of a policy x_t measured by the agent's **regret**

$$\operatorname{Reg}(T) = \max_{p \in \mathcal{X}} \int_0^T [u_t(p) - u_t(x_t)]\, dt = \max_{p \in \mathcal{X}} \int_0^T \langle v_t, p - x_t \rangle\, dt$$

No regret: $\operatorname{Reg}(T) = o(T)$ # the smaller the better

"The chosen policy is as good as the best fixed strategy in hindsight."

The agent's regret

Performance of a policy x_t measured by the agent's **regret**

$$\operatorname{Reg}(T) = \max_{p\in\mathcal{X}} \int_0^T [u_t(p) - u_t(x_t)]\,dt = \max_{p\in\mathcal{X}} \int_0^T \langle v_t, p - x_t\rangle\,dt$$

No regret: $\operatorname{Reg}(T) = o(T)$ # the smaller the better

"The chosen policy is as good as the best fixed strategy in hindsight."

Prolific literature:

- Economics ➺ Hannan (1957), Fudenberg & Levine (1998)
- Mathematics ➺ Blackwell (1956), Bubeck & Cesa-Bianchi (2012)
- Computer science ➺ Shalev-Shwartz (2011), Cesa-Bianchi & Lugosi (2006)

Exponential weights for online learning

Exponential weight dynamics

$$\dot{y}_t = v_t \qquad x_t = \Lambda(y_t) \tag{EWD}$$

where $\Lambda\colon \mathbb{R}^{\mathcal{A}} \to \mathcal{X}$ is the ***logit map***

$$\Lambda_a(y) = \frac{\exp(y_a)}{\sum_{a' \in \mathcal{A}} \exp(y_{a'})}$$

Does (EWD) lead to no regret?

Bounding the regret

- Fix a comparator $p \in \mathcal{X}$
- Consider associated regret

$$\mathrm{Reg}_p(T) = \int_0^T \langle v_t, p - x_t \rangle \, dt$$

Bounding the regret

- Fix a comparator $p \in \mathcal{X}$
- Consider associated regret

$$\mathrm{Reg}_p(T) = \int_0^T \langle v_t, p - x_t \rangle \, dt$$

- Focus on integrand

$$\langle v_t, x_t - p \rangle = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$

Bounding the regret

- Fix a comparator $p \in \mathcal{X}$
- Consider associated regret

$$\mathrm{Reg}_p(T) = \int_0^T \langle v_t, p - x_t \rangle \, dt$$

- Focus on integrand

$$\langle v_t, x_t - p \rangle = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$

- Suppose we can find a ***potential function*** $\Phi(y)$ such that

$$\nabla \Phi(y) = \Lambda(y) - p \implies \frac{d\Phi}{dt} = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$

Bounding the regret

- Fix a comparator $p \in \mathcal{X}$
- Consider associated regret
$$\mathrm{Reg}_p(T) = \int_0^T \langle v_t, p - x_t \rangle \, dt$$
- Focus on integrand
$$\langle v_t, x_t - p \rangle = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$
- Suppose we can find a ***potential function*** $\Phi(y)$ such that
$$\nabla \Phi(y) = \Lambda(y) - p \implies \frac{d\Phi}{dt} = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$
- Then
$$\mathrm{Reg}_p(T) = -\int_0^T \frac{d\Phi}{dt} \, dt = \Phi(y_0) - \Phi(y_T)$$

Bounding the regret

- Fix a comparator $p \in \mathcal{X}$
- Consider associated regret

$$\mathrm{Reg}_p(T) = \int_0^T \langle v_t, p - x_t \rangle \, dt$$

- Focus on integrand

$$\langle v_t, x_t - p \rangle = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$

- Suppose we can find a ***potential function*** $\Phi(y)$ such that

$$\nabla \Phi(y) = \Lambda(y) - p \implies \frac{d\Phi}{dt} = \langle \dot{y}_t, \Lambda(y_t) - p \rangle$$

- Then

$$\mathrm{Reg}_p(T) = -\int_0^T \frac{d\Phi}{dt} \, dt = \Phi(y_0) - \Phi(y_T)$$

If suitable potential exists $\implies \mathrm{Reg}(T) \le \Phi(y_0) - \min \Phi$

Finding a potential

What could a potential function look like?

Minimizing the potential

What is the minimum value of the potential?

Energy functions

We can encode the above with the help of the following ***energy functions:***

- **The Fenchel coupling:**

$$F(p, y) = \sum_{a \in \mathcal{A}} p_a \log p_a + \log \sum_{a \in \mathcal{A}} \exp(y_a) - \sum_{a \in \mathcal{A}} p_a y_a$$

- Substituting $x \leftarrow \Lambda(y)$ yields the **Kullback-Leibler divergence:**

$$D_{\mathrm{KL}}(p, x) = \sum_{a \in \mathcal{A}} p_a \log \frac{p_a}{x_a}$$

Key property: $\frac{d}{dt} F(p, y_t) = \langle v_t, x_t - p \rangle$

Regret of (EWD)

Theorem (Sorin (2009))

Under (EWD)*, the learner enjoys the regret bound*

$$\operatorname{Reg}_p(T) \leq F(p, y_0) = \sum_{a \in \mathcal{A}} p_a \log p_a + \log \sum_{a \in \mathcal{A}} \exp(y_{a,0}) - \sum_{a \in \mathcal{A}} p_a y_{a,0}$$

In particular, if (EWD) *is initialized with* $y_0 = 0$*, we have*

$$\operatorname{Reg}(T) \leq \log A$$

Online learning in discrete time

Sequence of events – discrete time

Require: set of actions $\mathcal{A}$; sequence of payoff vectors v_t, $t = 1, 2, \ldots$

for all $t = 1, 2, \ldots$ **do**

Choose **mixed strategy** $x_t \in \mathcal{X} := \Delta(\mathcal{A})$

Play **action** $a_t \sim x_t$

Encounter **payoff vector** v_t and receive **payoff** $u_t(a_t) = v_{a_t,t}$

end for

Defining elements

- Time: ***discrete***
- Players: single
- Actions: finite
- Payoffs: exogenous
- Feedback: ***depends*** (full or partial information, ...)

Online learning in discrete time

Sequence of events – discrete time

Require: set of actions $\mathcal{A}$; sequence of payoff vectors v_t, $t = 1, 2, \ldots$

for all $t = 1, 2, \ldots$ **do**

Choose **mixed strategy** $x_t \in \mathcal{X} := \Delta(\mathcal{A})$

Play **action** $a_t \sim x_t$

Encounter **payoff vector** v_t and receive **payoff** $u_t(a_t) = v_{a_t,t}$

end for

Regret

$$\mathrm{Reg}(T) = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} \left[\mathbb{E}_{v_{a_t,t}}[a_t \sim p] - \mathbb{E}_{v_{a_t,t}}[a_t \sim x_t] \right] = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} \langle v_t, p - x_t \rangle$$

The feedback process

Types of feedback

From best to worst (more to less info):

- ***Full information:*** v_t # deterministic vector feedback
- ***Noisy payoff vectors:*** $v_t + Z_t$ # stochastic vector feedback
- ***Bandit / Payoff-based:*** $u_t(a_t) = v_{a_t,t}$ # stochastic scalar feedback

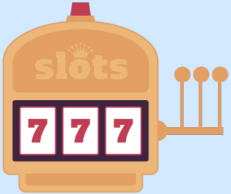
The feedback process

Types of feedback

From best to worst (more to less info):

- ***Full information:*** v_t — # deterministic vector feedback
- ***Noisy payoff vectors:*** $v_t + Z_t$ — # stochastic vector feedback
- ***Bandit / Payoff-based:*** $u_t(a_t) = v_{a_t,t}$ — # stochastic scalar feedback

Example



Play $x_t \leftarrow (1/2, 1/3, 1/6)$ $\rightsquigarrow$ Draw $a_t \leftarrow 1$

Full information

v_t 1 3 2

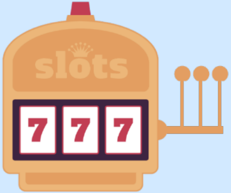
The feedback process

Types of feedback

From best to worst (more to less info):

- ***Full information:*** v_t # deterministic vector feedback
- ***Noisy payoff vectors:*** $v_t + Z_t$ # stochastic vector feedback
- ***Bandit / Payoff-based:*** $u_t(a_t) = v_{a_t,t}$ # stochastic scalar feedback

Example



Play $x_t \leftarrow (1/2, 1/3, 1/6)$ $\rightsquigarrow$ Draw $a_t \leftarrow 1$

Noisy payoff vectors

$v_t + Z_t$ 1.4 2.9 1.2

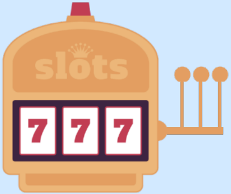
The feedback process

Types of feedback

From best to worst (more to less info):

- ***Full information:*** v_t # deterministic vector feedback
- ***Noisy payoff vectors:*** $v_t + Z_t$ # stochastic vector feedback
- ***Bandit / Payoff-based:*** $u_t(a_t) = v_{a_t,t}$ # stochastic scalar feedback

Example



Play $x_t \leftarrow (1/2, 1/3, 1/6)$ $\rightsquigarrow$ Draw $a_t \leftarrow 1$

Bandit / Payoff-based

$v_{a_t,t}$ 1 ✗ ✗

The feedback process

Types of feedback

From best to worst (more to less info):

- ***Full information:*** v_t # deterministic vector feedback
- ***Noisy payoff vectors:*** $v_t + Z_t$ # stochastic vector feedback
- ***Bandit / Payoff-based:*** $u_t(a_t) = v_{a_t,t}$ # stochastic scalar feedback

Defining features:

- ***Vector*** (all payoffs) vs. ***Scalar*** (bandit)
- ***Deterministic*** (full info) vs. ***Stochastic*** (noisy, bandit)

☞ Randomness defined relative to ***history of play*** $\mathcal{F}_t := \mathcal{F}(x_1, \dots, x_t)$

☞ Other feedback models also possible (noisy / delayed observations,...)

Regret

The agent's **regret** in discrete time

$$\textbf{Realized regret:} \quad \mathrm{Reg}(T) = \max_{a \in \mathcal{A}} \sum_{t=1}^{T} [u_t(a) - u_t(a_t)]$$

$$\textbf{Mean regret:} \quad \overline{\mathrm{Reg}}(T) = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} [u_t(p) - u_t(x_t)] = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} \langle v_t, p - x_t \rangle$$

Regret

The agent's **regret** in discrete time

Realized regret: $$\operatorname{Reg}(T) = \max_{a\in\mathcal{A}} \sum_{t=1}^{T} [u_t(a) - u_t(a_t)]$$

Mean regret: $$\overline{\operatorname{Reg}}(T) = \max_{p\in\mathcal{X}} \sum_{t=1}^{T} [u_t(p) - u_t(x_t)] = \max_{p\in\mathcal{X}} \sum_{t=1}^{T} \langle v_t, p - x_t\rangle$$

- **Adversarial framework:** regret guarantees against *any* given sequence v_t
- No distinction between ***mean*** regret and ***pseudo***-regret

➻ Bubeck & Cesa-Bianchi (2012)

- **Not here:** stochastic, Markovian, oblivious/non-oblivious,...

➻ Cesa-Bianchi & Lugosi (2006)

Feedback

Three types of feedback (from best to worst):

- **Full, exact information**: observe entire payoff vector v_t
- **Full, inexact information**: observe noisy estimate of v_t
- **Partial information / Bandit:** only chosen component $u_t(a_t) = v_{a_t,t}$

Feedback

Three types of feedback (from best to worst):

- **Full, exact information**: observe entire payoff vector v_t
- **Full, inexact information**: observe noisy estimate of v_t
- **Partial information / Bandit:** only chosen component $u_t(a_t) = v_{a_t,t}$

The oracle model

A ***stochastic first-order oracle (SFO)*** model of v_t is a random vector of the form

$$\hat{g}_t = v_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Assumptions

- ***Bias:*** $\|b_t\| \leq B_t$
- ***Variance:*** $\mathbb{E}[\|U_t\|^2 \mid \mathcal{F}_t] \leq \sigma_t^2$
- ***Second moment:*** $\mathbb{E}[\|\hat{g}_t\|^2 \mid \mathcal{F}_t] \leq M_t^2$

Reconstructing payoff vectors

Importance weighted estimators

Fix a payoff vector $v \in \mathbb{R}^{\mathcal{A}}$ and a probability distribution P on $\mathcal{A}$. Then the ***importance weighted estimator*** of v_a relative to P is the random variable

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

IWE as an oracle model

- ***Unbiased:***

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

- ***Second moment:***

$$\mathbb{E}[\hat{g}_a^2] = \frac{v_a^2}{P_a}$$

The Hedge algorithm

Algorithm HEDGE # ExpWeight with full information

Require: set of actions $\mathcal{A}$; sequence of payoff vectors $v_t \in [0,1]^{\mathcal{A}}$, $t = 1, 2, \ldots$

Initialize: $y_1 \in \mathbb{R}^{\mathcal{A}}$

for all $t = 1, 2, \ldots$ **do**

  set $x_t \leftarrow \Lambda(y_t)$ # mixed strategy

  play $a_t \sim x_t$ and **receive** $v_{a_t,t}$ # choose action / get payoff

  observe v_t # full info feedback

  set $y_{t+1} \leftarrow y_t + \gamma_t v_t$ # update scores

end for

Basic idea:

- Aggregate payoff information
- Choose actions with probability exponentially proportional to their scores
- Rinse & repeat

Regret analysis

- Use constant $\gamma_t \equiv \gamma$ # complications otherwise
- Fix benchmark strategy $p \in \mathcal{X}$ and consider the ***Fenchel coupling:***

$$F_t = F(p, y_t) = \sum_{a \in \mathcal{A}} p_a \log p_a + \log \sum_{a \in \mathcal{A}} \exp(y_{a,t}) - \langle y_t, p \rangle$$

- ***Energy inequality:***

$$F_{t+1} \leq F_t + \gamma \langle v_t, x_t - p \rangle + \tfrac{1}{2}\gamma^2 \|v_t\|_\infty^2$$

- Telescope to get

$$\mathrm{Reg}_p(T) \leq \frac{F_1}{\gamma} + \frac{\gamma T}{2}$$

- **How to proceed?**

Regret analysis, cont'd

How to choose γ?

Regret of Hedge

Theorem (Auer et al., 1995; Sorin, 2009)

☞ *Assume:*

- *sequence of payoff vectors* $v_t \in [0,1]^{\mathcal{A}}$; *full info feedback*
- $\gamma = \sqrt{(2\log A)/T}$

☞ *Then:* Hedge *enjoys the bound*

$$\mathrm{Reg}_p(T) \leq \sqrt{2\log A \cdot T} = \mathcal{O}(\sqrt{T})$$

Regret of Hedge

Theorem (Auer et al., 1995; Sorin, 2009)

Assume:

- *sequence of payoff vectors* $v_t \in [0,1]^{\mathcal{A}}$; *full info feedback*
- $\gamma = \sqrt{(2\log A)/T}$

Then: HEDGE *enjoys the bound*

$$\mathrm{Reg}_p(T) \leq \sqrt{2\log A \cdot T} = \mathcal{O}(\sqrt{T})$$

Remarks:

- Cannot achieve $\mathcal{O}(1)$ regret as in continuous time — # Why?
- This bound is tight in T — Abernethy et al., 2008
- Logarithmic dependence on A — Can deal with exponentially many arms!

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of v_t is a random vector $\hat{g}_t$ of the form

$$\hat{g}_t = v_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of v_t is a random vector $\hat{g}_t$ of the form

$$\hat{g}_t = v_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Assumptions

- ***Bias:*** $\|b_t\|_\infty \leq B_t$
- ***Variance:*** $\mathbb{E}[\|U_t\|_\infty^2 \mid \mathcal{F}_t] \leq \sigma_t^2$
- ***Second moment:*** $\mathbb{E}[\|\hat{g}_t\|_\infty^2 \mid \mathcal{F}_t] \leq M_t^2$

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of v_t is a random vector $\hat{g}_t$ of the form

$$\hat{g}_t = v_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Algorithm HEDGE-O — # ExpWeight with SFO feedback

Require: set of actions $\mathcal{A}$; sequence of payoff vectors $v_t \in \mathbb{R}^{\mathcal{A}}$, $t = 1, 2, \dots$

- **Initialize:** $y_1 \in \mathbb{R}^{\mathcal{A}}$
- **for all** $t = 1, 2, \dots$ **do**
  - **set** $x_t \leftarrow \Lambda(y_t)$ — # mixed strategy
  - **play** $a_t \sim x_t$ and **receive** $v_{a_t,t}$ — # choose action / get payoff
  - **observe** $\hat{g}_t \leftarrow v_t$ — # full info feedback
  - **set** $y_{t+1} \leftarrow y_t + \gamma_t \hat{g}_t$ — # update scores
- **end for**

Regret analysis

- Use constant $\gamma_t \equiv \gamma$ # complications otherwise
- Fix benchmark strategy $p \in \mathcal{X}$ and consider the ***Fenchel coupling:***

$$F_t = F(p, y_t) = \sum_{a \in \mathcal{A}} p_a \log p_a + \log \sum_{a \in \mathcal{A}} \exp(y_{a,t}) - \langle y_t, p \rangle$$

- ***Energy inequality:***

$$F_{t+1} \leq F_t + \gamma \langle \hat{g}_t, x_t - p \rangle + \tfrac{1}{2} \gamma^2 \|\hat{g}_t\|_\infty^2$$

- Expand and rearrange:

$$\langle v_t, p - x_t \rangle \leq \frac{F_t - F_{t+1}}{\gamma} + \langle U_t, x_t - p \rangle + \langle b_t, x_t - p \rangle + \frac{\gamma}{2} \|\hat{g}_t\|_\infty^2$$

- **How to proceed?**

Regret analysis, cont'd

Bound each term separately:

Regret of Hedge-O

Theorem

☞ **Assume:**

- *sequence of payoff vectors* $v_t \in \mathbb{R}^{\mathcal{A}}$; *SFO feedback*
- $\gamma = \sqrt{\dfrac{2 \log A}{\sum_{t=1}^{T} M_t^2}}$

☞ **Then:** *for all* $p \in \mathcal{X}$, Hedge-O *enjoys the bound*

$$\mathrm{Reg}_p(T) \leq 2 \sum_{t=1}^{T} B_t + \sqrt{2 \log A \cdot \sum_{t=1}^{T} M_t^2}$$

Regret of Hedge-O

Theorem

☞ ***Assume:***

- *sequence of payoff vectors $v_t \in \mathbb{R}^{\mathcal{A}}$; SFO feedback*
- $\gamma = \sqrt{\dfrac{2\log A}{\sum_{t=1}^{T} M_t^2}}$

☞ ***Then:*** *for all $p \in \mathcal{X}$,* HEDGE-O *enjoys the bound*

$$\operatorname{Reg}_p(T) \le 2\sum_{t=1}^{T} B_t + \sqrt{2\log A \cdot \sum_{t=1}^{T} M_t^2}$$

Remarks:

- $\mathcal{O}(\sqrt{T})$ regret if feedback is unbiased ($b_t = 0$) and has finite variance ($M_t \le M$)
- This bound is tight in T — Abernethy et al., 2008
- Logarithmic dependence on A — Can deal with exponentially many arms!

Learning with bandit feedback

Three types of feedback (from best to worst):

- **Full, exact information:** observe entire payoff vector v_t
- **Full, inexact information:** observe noisy estimate of v_t
- **Partial information / Bandit:** only chosen component $u_t(a_t) = v_{a_t,t}$

Importance weighted estimators

Fix a payoff vector $v \in \mathbb{R}^{\mathcal{A}}$ and a probability distribution P on $\mathcal{A}$. Then the ***importance weighted estimator*** of v_a is the random variable

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

IWE as an oracle model

- ***Unbiased:*** $\mathbb{E}[\hat{g}_a] = v_a$ ☞ $b_t = 0$
- ***Second moment:*** $\mathbb{E}[\hat{g}_a^2] = v_a^2 / P_a$ ☞ $M_t = \mathcal{O}(1/\min_a x_{a,t})$

The EXP3 algorithm

Algorithm Exponential weights for exploration and exploitation (EXP3) # Hedge with bandit feedback

Require: set of actions $\mathcal{A}$; sequence of payoff vectors $v_t \in [0,1]^{\mathcal{A}}$, $t = 1, 2, \ldots$

Initialize: $y_1 \in \mathbb{R}^{\mathcal{A}}$

for all $t = 1, 2, \ldots$ **do**

set $x_t \leftarrow \Lambda(y_t)$ # mixed strategy

play $a_t \sim x_t$ and **receive** $v_{a_t,t}$ # choose action / get payoff

set $\hat{g}_t \leftarrow \dfrac{v_{a_t,t}}{x_{a_t,t}} e_{a_t}$ # IW estimator

set $y_{t+1} \leftarrow y_t + \gamma_t \hat{g}_t$ # update scores

end for

Regret analysis

- Use constant $\gamma_t \equiv \gamma$ # complications otherwise
- Fix benchmark strategy $p \in \mathcal{X}$ and consider the ***Fenchel coupling:***

$$F_t = F(p, y_t) = \sum_{a \in \mathcal{A}} p_a \log p_a + \log \sum_{a \in \mathcal{A}} \exp(y_{a,t}) - \langle y_t, p \rangle$$

- ***Energy inequality:***

$$F_{t+1} \leq F_t + \gamma \langle \hat{g}_t, x_t - p \rangle + \tfrac{1}{2}\gamma^2 \|\hat{g}_t\|_\infty^2$$

- Expand and rearrange:

$$\langle v_t, p - x_t \rangle \leq \frac{F_t - F_{t+1}}{\gamma} + \langle U_t, x_t - p \rangle + \langle b_t, x_t - p \rangle + \frac{\gamma}{2} \|\hat{g}_t\|_\infty^2$$

- **How to proceed?**

Energy inequality

Basic lemma

Fix some $y, w \in \mathbb{R}^{\mathcal{A}}$, and let $x \propto \exp(y)$. Then:

$$\log \sum_{a\in\mathcal{A}} \exp(y_a + w_a) \le \log \sum_{a\in\mathcal{A}} \exp(y_a) + \langle x, w\rangle + \tfrac{1}{2}\|w\|_\infty^2$$

Energy inequality

Basic lemma

Fix some $y \in \mathbb{R}^{\mathcal{A}}$, $w \in (-\infty, 1]^{\mathcal{A}}$, and let $x \propto \exp(y)$. Then:

$$\log \sum_{a \in \mathcal{A}} \exp(y_a + w_a) \leq \log \sum_{a \in \mathcal{A}} \exp(y_a) + \langle x, w \rangle + \sum_{a \in \mathcal{A}} x_a w_a^2$$

Proof.

□

Regret analysis, cont'd

Regret of EXP3

Theorem (Auer et al., 1995)

☞ *Assume:*

- *EXP3 is run for T iterations with $\gamma = \sqrt{\log A/(AT)}$*

- **Then:** *For all $p \in \mathcal{X}$, the learner enjoys the bound*

$$\mathbb{E}[\mathrm{Reg}_p(T)] \leq 2\sqrt{A \log A \cdot T}$$

Regret of EXP3

Theorem (Auer et al., 1995)

☞ *Assume:*

- *EXP3 is run for T iterations with $\gamma = \sqrt{\log A/(AT)}$*

- *Then: For all $p \in \mathcal{X}$, the learner enjoys the bound*

$$\mathbb{E}[\mathrm{Reg}_p(T)] \leq 2\sqrt{A \log A \cdot T}$$

Remarks:

- ✓ Tight in T — Abernethy et al., 2008
- ✗ Worse than full info bound by a factor of $\sqrt{A}$ — # cf. Hedge-O
- Regret can be improved to $\mathcal{O}(\sqrt{AT})$ **but no lower** — Audibert & Bubeck, 2010; Abernethy et al., 2015
- T must be known — Thoughts?
- (IWE) is still unbounded — Thoughts?

Outline

Setting

Sequence of events: Online convex optimization (OCO)

Require: convex action set $\mathcal{X} \subseteq \mathbb{R}^d$; convex loss functions $\ell_t\colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

repeat

At each epoch $t = 1, 2, \dots$ **do**

Choose ***action*** $x_t \in \mathcal{X}$ # action selection

Encounter ***loss function*** $\ell_t\colon \mathcal{X} \to \mathbb{R}$ # Nature plays

Incur ***cost*** $c_t = \ell_t(x_t)$ # reward phase

Observe loss function ℓ_t # feedback phase

until end

Defining elements

- ***Time:*** discrete
- ***Players:*** single
- ***Actions:*** continuous
- ***Losses:*** exogenous
- ***Feedback:*** **depends** (function-based, gradient-based, loss-based, ...)

Setting

Sequence of events: Online convex optimization (OCO)

Require: convex **action set** $\mathcal{X} \subseteq \mathbb{R}^d$; convex **loss functions** $\ell_t \colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

repeat

At each epoch $t = 1, 2, \dots$ **do**

Choose ***action*** $x_t \in \mathcal{X}$ — # action selection

Encounter ***loss function*** $\ell_t \colon \mathcal{X} \to \mathbb{R}$ — # Nature plays

Incur ***cost*** $c_t = \ell_t(x_t)$ — # reward phase

Observe ***gradient*** $g_t = \nabla \ell_t(x_t)$ — # feedback phase

until end

Defining elements

- ***Time:*** discrete
- ***Players:*** single
- ***Actions:*** continuous
- ***Losses:*** exogenous
- ***Feedback:*** **depends** (function-based, ***gradient-based,*** loss-based, ...)

Setting

Sequence of events: Online convex optimization (OCO)

Require: convex **action set** $\mathcal{X} \subseteq \mathbb{R}^d$; convex **loss functions** $\ell_t \colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

repeat

At each epoch $t = 1, 2, \dots$ **do**

Choose ***action*** $x_t \in \mathcal{X}$ — # action selection

Encounter ***loss function*** $\ell_t \colon \mathcal{X} \to \mathbb{R}$ — # Nature plays

Incur ***cost*** $c_t = \ell_t(x_t)$ — # reward phase

Observe **cost** $c_t = \ell_t(x_t)$ — # feedback phase

until end

Defining elements

- ***Time:*** discrete
- ***Players:*** single
- ***Actions:*** continuous
- ***Losses:*** exogenous
- ***Feedback:*** **depends** (function-based, gradient-based, **loss-based,** ...)

Feedback

Types of feedback

From best to worst (more to less info):

- ***Full information:*** observe entire loss function $\ell_t \colon \mathcal{X} \to \mathbb{R}$ # deterministic function feedback
- ***First-order info, exact:*** observe (sub)gradient $g_t \in \partial \ell_t(x_t)$ # deterministic vector feedback
- ***First-order info, inexact***: observe noisy estimate of g_t # stochastic vector feedback
- ***Zeroth-order info (bandit):*** observe only incurred cost $c_t = \ell_t(x_t)$ # deterministic scalar feedback

Feedback

Types of feedback

From best to worst (more to less info):

- *Full information:* observe entire loss function $\ell_t \colon \mathcal{X} \to \mathbb{R}$ # deterministic function feedback
- *First-order info, exact:* observe (sub)gradient $g_t \in \partial \ell_t(x_t)$ # deterministic vector feedback
- *First-order info, inexact*: observe noisy estimate of g_t # stochastic vector feedback
- *Zeroth-order info (bandit):* observe only incurred cost $c_t = \ell_t(x_t)$ # deterministic scalar feedback

The oracle model

A *stochastic first-order oracle (SFO)* for $g_t \in \partial \ell_t(x_t)$ is a random vector of the form

$$\hat{g}_t = g_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - g_t$ is the **bias** of $\hat{g}_t$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$[\ell_t(x_t) - \ell_t(p)]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\operatorname{Reg}(T) = \max_{p\in\mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p\in\mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\operatorname{Reg}(T) = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

- ***No regret:*** $\operatorname{Reg}(T) = o(T)$
- ***Adversarial framework:*** minimize regret against **any** given sequence ℓ_t

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\mathrm{Reg}(T) = \max_{p\in\mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p\in\mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

- ***No regret:*** $\mathrm{Reg}(T) = o(T)$
- ***Adversarial framework:*** minimize regret against **any** given sequence ℓ_t
- ***Expected regret:***

$$\mathbb{E}[\mathrm{Reg}(T)] = \mathbb{E}\left[\max_{p\in\mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

- ***Pseudo-regret:***

$$\overline{\mathrm{Reg}}(T) = \max_{p\in\mathcal{X}} \mathbb{E}\left[\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\mathrm{Reg}(T) = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

- ***No regret:*** $\mathrm{Reg}(T) = o(T)$
- ***Adversarial framework:*** minimize regret against **any** given sequence ℓ_t
- ***Expected regret:***

$$\mathbb{E}[\mathrm{Reg}(T)] = \mathbb{E}\left[\max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

- ***Pseudo-regret:***

$$\overline{\mathrm{Reg}}(T) = \max_{p \in \mathcal{X}} \mathbb{E}\left[\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

- $\overline{\mathrm{Reg}}(T) \le \mathbb{E}[\mathrm{Reg}(T)]$: bounds do not translate "as is" but "almost"

➡ Cesa-Bianchi & Lugosi, 2006, Bubeck & Cesa-Bianchi, 2012, Lattimore & Szepesvári, 2020

Be the leader

- Suppose ℓ_t is observed ***before*** playing x_t
- Then the agent can try to ***be the leader (BTL)***

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{BTL}$$

Be the leader

- Suppose ℓ_t is observed ***before*** playing x_t
- Then the agent can try to ***be the leader (BTL)***

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{BTL}$$

Regret of BTL

☞ Under (BTL), the learner incurs $\mathrm{Reg}(T) = 0$.

Be the leader

- Suppose ℓ_t is observed ***before*** playing x_t
- Then the agent can try to ***be the leader (BTL)***

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{BTL}$$

Regret of BTL

☞ Under (BTL), the learner incurs $\mathrm{Reg}(T) = 0$.

...unrealistic

Follow the leader

- Suppose ℓ_t is observed ***after*** playing x_t
- Then the agent can try to ***follow the leader (FTL)***

$$x_{t+1} \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{FTL}$$

Follow the leader

- Suppose ℓ_t is observed ***after*** playing x_t
- Then the agent can try to ***follow the leader (FTL)***

$$x_{t+1} \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{FTL}$$

Does (FTL) lead to no regret?

Template bound for FTL

FTL regret bound

For all $p \in \mathcal{X}$, the regret of (FTL) can be bounded as

$$\mathrm{Reg}_p(T) = \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] \leq \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(x_{t+1})]$$

Template bound for FTL

FTL regret bound

For all $p \in \mathcal{X}$, the regret of (FTL) can be bounded as

$$\mathrm{Reg}_p(T) = \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] \leq \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(x_{t+1})]$$

Proof.

□

FTL against quadratic losses

Test (FTL) in an ***online quadratic optimization (OQO)*** problem:

$$\ell_t(x) = \frac{1}{2}\|x - p_t\|^2 \quad \text{for some sequence of center points } p_t,\ t = 1, 2, \dots \tag{OQO}$$

FTL against quadratic losses

Test (FTL) in an ***online quadratic optimization (OQO)*** problem:

$$\ell_t(x) = \frac{1}{2}\|x - p_t\|^2 \quad \text{for some sequence of center points } p_t,\, t = 1, 2, \ldots \tag{OQO}$$

Regret of FTL in quadratic problems

☞ ***Assume:*** (FTL) is run against (OQO) with $\sup_t \|p_t\| \leq R$

✓ **Then:** $\operatorname{Reg}(T) \leq 4R^2(1 + \log T)$

FTL against quadratic losses

Test (FTL) in an ***online quadratic optimization (OQO)*** problem:

$$\ell_t(x) = \frac{1}{2}\|x - p_t\|^2 \quad \text{for some sequence of center points } p_t,\ t = 1, 2, \ldots \tag{OQO}$$

Regret of FTL in quadratic problems

☞ ***Assume:*** (FTL) is run against (OQO) with $\sup_t \|p_t\| \leq R$

✓ **Then:** $\operatorname{Reg}(T) \leq 4R^2(1 + \log T)$

Proof.

□

FTL against linear losses

Test (FTL) in an ***online linear optimization (OLO)*** problem:

$$\ell_t(x) = \langle w_t, x \rangle \quad \text{for some sequence of loss vectors } w_t \in \mathbb{R}^d,\ t = 1, 2, \dots \tag{OLO}$$

FTL against linear losses

Test (FTL) in an ***online linear optimization (OLO)*** problem:

$$\ell_t(x) = \langle w_t, x \rangle \quad \text{for some sequence of loss vectors } w_t \in \mathbb{R}^d,\ t = 1, 2, \dots \tag{OLO}$$

Chasing the leader

☞ ***Assume:*** $\mathcal{X} = [-1, 1]$ and (FTL) is run against (OLO) with $w_1 = -1/2$ and $w_t = (-1)^t$ otherwise

⚠ ***What is the incurred regret?***

Follow the regularized leader

Add a fictitious "day zero loss" $\Longrightarrow$ ***follow the regularized leader (FTRL)***

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \Big\{ \sum_{s=1}^{t} \ell_s(x) + \underbrace{\lambda h(x)}_{\text{“}\ell_0(x)\text{”}} \Big\} \tag{FTRL}$$

where

- The ***regularization function*** $h\colon \mathcal{X} \to \mathbb{R}$ is strongly convex # $h(x) - (K/2)\|x\|^2$ convex for some $K > 0$
- The ***regularization weight*** $\lambda > 0$ can be tuned by the optimizer

Main idea: Regularization $\Longrightarrow$ Stability $\Longrightarrow$ Less regret

➺ Algorithm due to Shalev-Shwartz & Singer, 2006, Shalev-Shwartz, 2011

Example 1: Euclidean regularization

- **Setup:** $\mathcal{X} = \mathbb{R}^d$, linear losses $\ell_t(x) = \langle w_t, x \rangle$
- **Regularizer:**
$$h(x) = \tfrac{1}{2}\|x\|^2$$
- **Algorithm:**
$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle w_s, x \rangle + \frac{\lambda}{2}\|x\|^2 \right\}$$

Example 1: Euclidean regularization

- **Setup:** $\mathcal{X} = \mathbb{R}^d$, linear losses $\ell_t(x) = \langle w_t, x \rangle$
- **Regularizer:**

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

- **Algorithm:**

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle w_s, x \rangle + \frac{\lambda}{2} \|x\|^2 \right\} = -\frac{1}{\lambda} \sum_{s=1}^{t} w_s = x_t - (1/\lambda) w_t$$

Example 1: Euclidean regularization

- **Setup:** $\mathcal{X} = \mathbb{R}^d$, linear losses $\ell_t(x) = \langle w_t, x \rangle$
- **Regularizer:**
$$h(x) = \tfrac{1}{2}\|x\|^2$$
- **Algorithm:**
$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle w_s, x \rangle + \frac{\lambda}{2}\|x\|^2 \right\} = -\frac{1}{\lambda}\sum_{s=1}^{t} w_s = x_t - (1/\lambda) w_t$$
- Euclidean regularization + linear losses $(w_t = \nabla \ell_t(x_t)) \implies$ ***gradient descent:***
$$x_{t+1} = x_t - \underbrace{\eta}_{1/\lambda} \nabla \ell_t(x_t) \tag{GD}$$

Example 2: Entropic regularization

- **Setup:** $\mathcal{X} = \Delta(\mathcal{A})$, linear payoffs $u_t(x) = \langle v_t, x \rangle$

☞ payoffs instead of costs

- **Regularizer:**

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

- **Algorithm:**

$$x_{t+1} = \arg\max_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle v_s, x \rangle - \lambda \sum_{a \in \mathcal{A}} x_a \log x_a \right\}$$

Example 2: Entropic regularization

- **Setup:** $\mathcal{X} = \Delta(\mathcal{A})$, linear payoffs $u_t(x) = \langle v_t, x \rangle$

☞ payoffs instead of costs

- **Regularizer:**

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

- **Algorithm:**

$$x_{t+1} = \arg\max_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle v_s, x \rangle - \lambda \sum_{a \in \mathcal{A}} x_a \log x_a \right\} = \frac{\exp(\sum_{s=1}^{t} v_{a,s}/\lambda)}{\sum_{a' \in \mathcal{A}} \exp(\sum_{s=1}^{t} v_{a',s}/\lambda)}$$

Example 2: Entropic regularization

- **Setup:** $\mathcal{X} = \Delta(\mathcal{A})$, linear payoffs $u_t(x) = \langle v_t, x\rangle$ ☞ payoffs instead of costs
- **Regularizer:**

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

- **Algorithm:**

$$x_{t+1} = \arg\max_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle v_s, x\rangle - \lambda \sum_{a \in \mathcal{A}} x_a \log x_a \right\} = \frac{\exp(\sum_{s=1}^{t} v_{a,s}/\lambda)}{\sum_{a' \in \mathcal{A}} \exp(\sum_{s=1}^{t} v_{a',s}/\lambda)}$$

- Entropic regularization + linear payoffs $\implies$ ***exponential weights:***

$$\begin{aligned} y_{t+1} &= y_t + \overbrace{\eta}^{1/\lambda} v_t \\ x_{t+1} &= \underbrace{\Lambda(y_{t+1})}_{\text{logit map}} \end{aligned} \tag{EW}$$

Template bound for FTRL

FTRL regret bound

For all $p \in \mathcal{X}$, the regret of (FTRL) can be bounded as

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - h(x_1)] + \sum_{t=1}^{T}[\ell_t(x_t) - \ell_t(x_{t+1})]$$

Template bound for FTRL

FTRL regret bound

For all $p \in \mathcal{X}$, the regret of (FTRL) can be bounded as

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - h(x_1)] + \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(x_{t+1})]$$

Proof.

□

Variability bound for FTRL

Variability of FTRL

☞ ***Assume:*** h is K-strongly convex; each ℓ_t is G_t-Lipschitz continuous

✓ **Then:**

$$\ell_t(x_t) - \ell_t(x_{t+1}) \leq G_t \|x_{t+1} - x_t\| \leq G_t^2/(\lambda K)$$

Variability bound for FTRL

Variability of FTRL

- ☞ ***Assume:*** h is K-strongly convex; each ℓ_t is G_t-Lipschitz continuous
- ✓ **Then:**

$$\ell_t(x_t) - \ell_t(x_{t+1}) \leq G_t \|x_{t+1} - x_t\| \leq G_t^2/(\lambda K)$$

Proof.

□

Regret of FTRL

Theorem (Shalev-Shwartz & Singer, 2006; Shalev-Shwartz, 2011)

- **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous*
- ✓ ***Then:*** (FTRL) *enjoys the regret bound*

$$\operatorname{Reg}_p(T) \leq \lambda[h(p) - \min h] + \frac{G^2}{\lambda K} T$$

Regret of FTRL

Theorem (Shalev-Shwartz & Singer, 2006; Shalev-Shwartz, 2011)

- **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous*
- ✓ **Then:** (FTRL) *enjoys the regret bound*

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - \min h] + \frac{G^2}{\lambda K} T$$

Corollary

With assumptions as above, $H = \max h - \min h$ and $\lambda = G\sqrt{T/(2KH)}$, (FTRL) *enjoys the bound*

$$\mathrm{Reg}(T) \leq G\sqrt{(2H/K)\,T} = \mathcal{O}(\sqrt{T})$$

Regret of FTRL

Theorem (Shalev-Shwartz & Singer, 2006; Shalev-Shwartz, 2011)

☞ **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous*

✓ ***Then:*** (FTRL) *enjoys the regret bound*

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - \min h] + \frac{G^2}{\lambda K} T$$

Corollary

With assumptions as above, $H = \max h - \min h$ and $\lambda = G\sqrt{T/(2KH)}$, (FTRL) *enjoys the bound*

$$\mathrm{Reg}(T) \leq G\sqrt{(2H/K)\,T} = \mathcal{O}(\sqrt{T})$$

Remarks:

- The bound is tight in T — Abernethy et al., 2008
- Requires full information and tuning in terms of T — # can relax

Feedback

Types of feedback

From best to worst (more to less info):

- ***Full information:*** observe entire loss function $\ell_t: \mathcal{X} \to \mathbb{R}$ # deterministic function feedback
- ***First-order info, exact:*** observe (sub)gradient $g_t \in \partial \ell_t(x_t)$ # deterministic vector feedback
- ***First-order info, inexact***: observe noisy estimate of g_t # stochastic vector feedback
- ***Zeroth-order info (bandit):*** observe only incurred cost $c_t = \ell_t(x_t)$ # deterministic scalar feedback

Feedback

Types of feedback

From best to worst (more to less info):

- *Full information:* observe entire loss function $\ell_t \colon \mathcal{X} \to \mathbb{R}$ # deterministic function feedback
- *First-order info, exact:* observe (sub)gradient $g_t \in \partial \ell_t(x_t)$ # deterministic vector feedback
- *First-order info, inexact*: observe noisy estimate of g_t # stochastic vector feedback
- *Zeroth-order info (bandit):* observe only incurred cost $c_t = \ell_t(x_t)$ # deterministic scalar feedback

The oracle model

A *stochastic first-order oracle (SFO)* for $g_t \in \partial \ell_t(x_t)$ is a random vector of the form

$$\hat{g}_t = g_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Follow the linearized leader

Can we relax the full information requirement of FTRL?

- Replace ℓ_t with first-order surrogate

$$\hat{\ell}_t(x) = \ell_t(x_t) + \langle g_t, x - x_t \rangle \qquad g_t \in \partial \ell_t(x_t)$$

- Plug into (FTRL)

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \hat{\ell}_s(x) + \underbrace{\lambda}_{1/\eta} h(x) \right\} = \arg\min_{x \in \mathcal{X}} \left\{ \eta \sum_{s=1}^{t} \langle g_s, x - x_s \rangle + h(x) \right\}$$

Follow the linearized leader

Can we relax the full information requirement of FTRL?

- Replace ℓ_t with first-order surrogate

$$\hat{\ell}_t(x) = \ell_t(x_t) + \langle g_t, x - x_t \rangle \qquad g_t \in \partial \ell_t(x_t)$$

- Plug into (FTRL)

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \hat{\ell}_s(x) + \underbrace{\lambda}_{1/\eta} h(x) \right\} = \arg\min_{x \in \mathcal{X}} \left\{ \eta \sum_{s=1}^{t} \langle g_s, x - x_s \rangle + h(x) \right\}$$

- ***Follow the linearized leader (FTLL)***

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \eta \sum_{s=1}^{t} \langle g_s, x \rangle + h(x) \right\} \tag{FTLL}$$

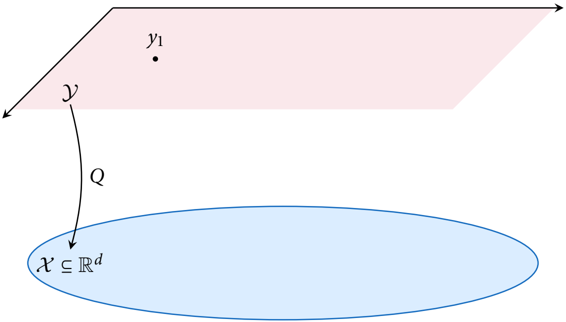
Dual averaging

Dual averaging (DA) formulation of FTLL

Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h

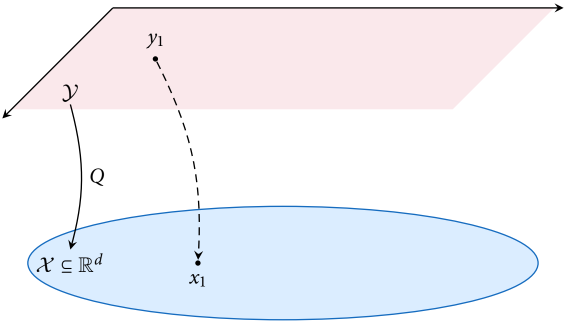


Dual averaging

Dual averaging (DA) formulation of FTLL ➺ Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***mirror map*** associated to h



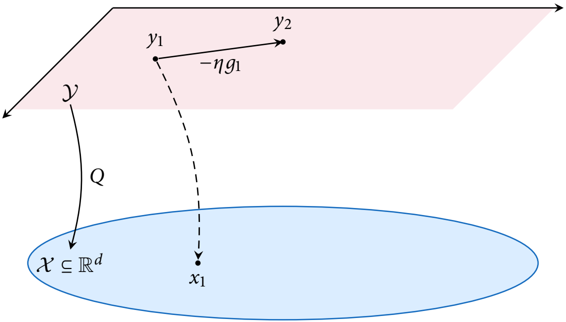
Dual averaging

Dual averaging (DA) formulation of FTLL

Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***mirror map*** associated to h

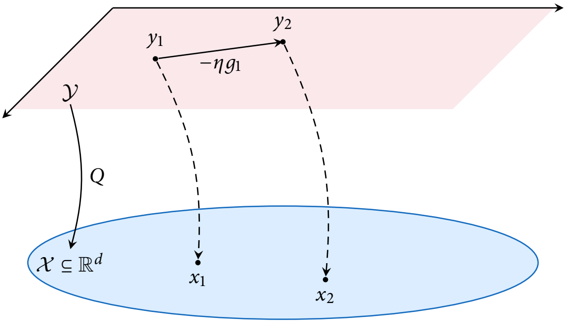


Dual averaging

Dual averaging (DA) formulation of FTLL — Nesterov, 2009; Xiao, 2010

$$
\begin{aligned}
y_{t+1} &= y_t - \eta g_t \\
x_{t+1} &= Q(y_{t+1})
\end{aligned}
\tag{DA}
$$

where $Q(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h

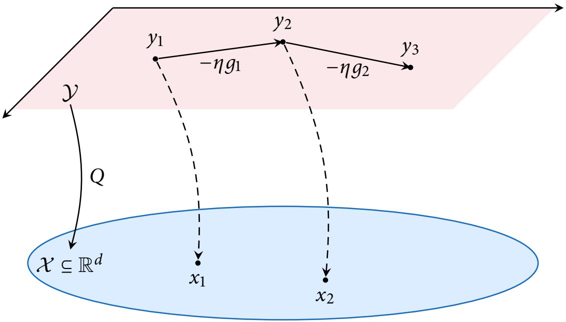


Dual averaging

Dual averaging (DA) formulation of FTLL — Nesterov, 2009; Xiao, 2010

$$
\begin{aligned}
y_{t+1} &= y_t - \eta g_t \\
x_{t+1} &= Q(y_{t+1})
\end{aligned}
\tag{DA}
$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h

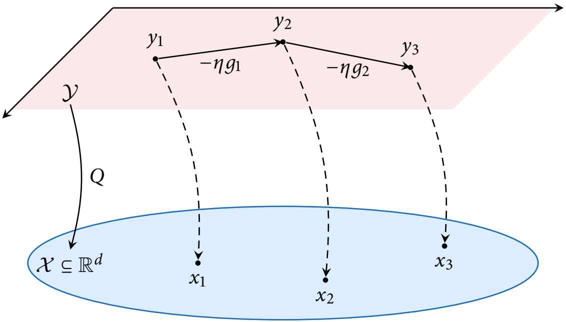


Dual averaging

Dual averaging (DA) formulation of FTLL → Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x\in\mathcal{X}} \{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h



Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

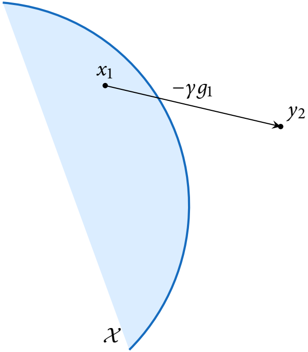


Figure: Schematics of (OGD)

Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

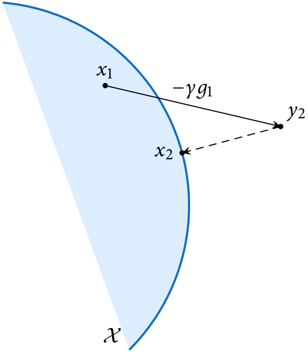


Figure: Schematics of (OGD)

Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

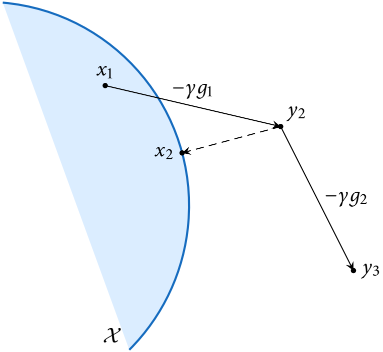


Figure: Schematics of (OGD)

Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

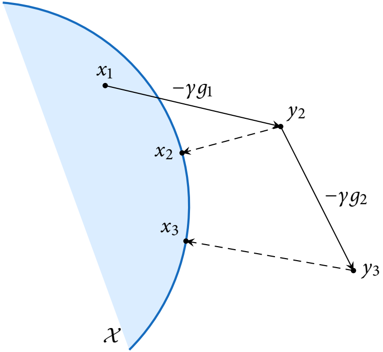


Figure: Schematics of (OGD)

Online mirror descent (deep dive)

- Gradient signals enter (DA) unweighted / unadjusted # post-adaptation
- Variable weights $\rightsquigarrow$ "lazy", primal-dual variant of **online mirror descent**

$$\begin{aligned} y_{t+1} &= y_t + \eta_t \hat{g}_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \qquad (\text{OMD}_{\text{lazy}})$$

- Primal-primal ("eager") variant of $(\text{OMD}_{\text{lazy}})$

$$x_{t+1} = P_{x_t}(\eta_t \hat{g}_t) \qquad (\text{OMD})$$

with the **Bregman proximal mapping** P defined as

$$P_x(w) = \arg\min_{x' \in \mathcal{X}} \{\langle w, x - x' \rangle + D(x', x)\}$$

where $D(x', x) = h(x') - h(x) - \langle \nabla h(x'), x - x' \rangle$ is the **Bregman divergence** of h

Online mirror descent (deep dive)

- Gradient signals enter (DA) unweighted / unadjusted # post-adaptation
- Variable weights $\rightsquigarrow$ "lazy", primal-dual variant of **online mirror descent**

$$\begin{aligned} y_{t+1} &= y_t + \eta_t \hat{g}_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{OMD$_{\text{lazy}}$}$$

- Primal-primal ("eager") variant of (OMD$_{\text{lazy}}$)

$$x_{t+1} = P_{x_t}(\eta_t \hat{g}_t) \tag{OMD}$$

with the **Bregman proximal mapping** P defined as

$$P_x(w) = \arg\min_{x' \in \mathcal{X}} \{ \langle w, x - x' \rangle + D(x', x) \}$$

where $D(x', x) = h(x') - h(x) - \langle \nabla h(x'), x - x' \rangle$ is the **Bregman divergence** of h

Proposition

The iterates of (OMD$_{\text{lazy}}$) *and* (OMD) *coincide whenever* $\operatorname{dom} \partial h = \operatorname{ri} \mathcal{X}$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- ***Energy function:*** ⚠ take for granted

$$F_t = h(p) + h^*(y_t) - \langle y_t, p \rangle$$

where $h^*(y) = \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***potential*** of $Q \rightsquigarrow \nabla h^* = Q$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- ***Energy function:*** ⚠ take for granted

$$F_t = h(p) + h^*(y_t) - \langle y_t, p \rangle$$

where $h^*(y) = \max_{x \in \mathcal{X}} \{ \langle y, x \rangle - h(x) \}$ is the ***potential*** of $Q \rightsquigarrow \nabla h^* = Q$

- ***Template inequality:*** ⚠ take for granted

$$F_{t+1} \leq F_t - \eta \langle g_t, x_t - p \rangle + \frac{\eta^2}{2K} \|g_t\|^2$$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- ***Energy function:*** ⚠ take for granted

$$F_t = h(p) + h^*(y_t) - \langle y_t, p \rangle$$

where $h^*(y) = \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the **potential** of $Q \rightsquigarrow \nabla h^* = Q$

- ***Template inequality:*** ⚠ take for granted

$$F_{t+1} \leq F_t - \eta \langle g_t, x_t - p \rangle + \frac{\eta^2}{2K} \|g_t\|^2$$

- ***Rearrange & telescope:*** # build the regret

$$\overline{\mathrm{Reg}}(T) \leq \frac{H}{\eta} + \frac{\eta}{2K} \sum_{t=1}^{T} G_t^2$$

Regret under dual averaging, cont'd

- Take $\eta = \sqrt{2KH / \sum_{t=1}^{T} G_t^2}$

⚠ Why?

$$\operatorname{Reg}(T) \leq \sqrt{(2H/K) \sum_{t=1}^{T} G_t^2}$$

Regret under dual averaging, cont'd

- Take $\eta = \sqrt{2KH / \sum_{t=1}^{T} G_t^2}$ ⚠ Why?

$$\operatorname{Reg}(T) \leq \sqrt{(2H/K) \sum_{t=1}^{T} G_t^2}$$

Theorem (Shalev-Shwartz, 2011)

☞ **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous; $H = \max h - \min h$ and $\eta = G^{-1}\sqrt{2KH/T}$*

✓ ***Then:*** (DA) / (FTLL) *enjoys the regret bound*

$$\operatorname{Reg}_p(T) \leq G\sqrt{(2H/K)T}$$

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of g_t is a random vector $\hat{g}_t$ of the form

$$\hat{g}_t = g_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of g_t is a random vector $\hat{g}_t$ of the form

$$\hat{g}_t = g_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Assumptions

- ***Bias:*** $\|b_t\|_\infty \le B_t$
- ***Variance:*** $\mathbb{E}[\|U_t\|_\infty^2 \mid \mathcal{F}_t] \le \sigma_t^2$
- ***Second moment:*** $\mathbb{E}[\|\hat{g}_t\|_\infty^2 \mid \mathcal{F}_t] \le M_t^2$

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of g_t is a random vector $\hat{g}_t$ of the form

$$\hat{g}_t = g_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Algorithm Stochastic gradient descent (SGD) # OGD with stochastic feedback

Require: convex **action set** $\mathcal{X} \subseteq \mathbb{R}^d$; convex **loss functions** $\ell_t : \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

Initialize: $y_1 \in \mathbb{R}^{\mathcal{A}}$

for all $t = 1, 2, \dots$ **do**

- **play** $x_t \leftarrow \Pi(y_t)$ # action selection
- **incur** $c_t = \ell_t(x_t)$ # incur cost
- **observe** estimate $\hat{g}_t$ of $g_t \in \partial \ell_t(x_t)$ # SFO feedback
- **set** $y_{t+1} \leftarrow y_t - \eta_t \hat{g}_t$ # update state

end for

Regret under OGD

- **Gradient trick:** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p\rangle \quad \text{for all } p \in \mathcal{X}$$

- **Energy function:** # as before

$$F_t = \tfrac{1}{2}\|y_t - p\|^2 - \tfrac{1}{2}\|y_t - x_t\|^2$$

- **Energy inequality:** # $\hat{g}_t$ instead of g_t

$$F_{t+1} \leq F_t - \eta\langle \hat{g}_t, x_t - p\rangle + \frac{\eta^2}{2}\|\hat{g}_t\|^2$$

- **Expand and rearrange:**

$$\langle v_t, p - x_t\rangle \leq \frac{F_t - F_{t+1}}{\eta} - \langle U_t, x_t - p\rangle - \langle b_t, x_t - p\rangle + \frac{\eta}{2}\|\hat{g}_t\|_\infty^2$$

- **How to proceed?**

Regret analysis, cont'd

Bound each term separately:

Regret of SGD

Theorem

☞ ***Assume:***

- *feedback of the form* (SFO)
- $\eta = \operatorname{diam}(\mathcal{X}) \Big/ \sqrt{\sum_{t=1}^{T} M_t^2}$

✓ ***Then:*** *for all $p \in \mathcal{X}$, the SGD algorithm enjoys the bound*

$$\mathbb{E}[\operatorname{Reg}_p(T)] \leq 2 \sum_{t=1}^{T} B_t + \operatorname{diam}(\mathcal{X}) \sqrt{\sum_{t=1}^{T} M_t^2}$$

Regret of SGD

Theorem

☞ ***Assume:***

- *feedback of the form* (SFO)
- $\eta = \operatorname{diam}(\mathcal{X}) \Big/ \sqrt{\sum_{t=1}^{T} M_t^2}$

✓ ***Then:*** *for all $p \in \mathcal{X}$, the SGD algorithm enjoys the bound*

$$\mathbb{E}[\operatorname{Reg}_p(T)] \le 2\sum_{t=1}^{T} B_t + \operatorname{diam}(\mathcal{X})\sqrt{\sum_{t=1}^{T} M_t^2}$$

Remarks:

- $\mathcal{O}(\sqrt{T})$ regret if feedback is unbiased ($b_t = 0$) and has finite variance ($M_t \le M$)
- This bound is tight in T ➺ Abernethy et al., 2008

Stochastic convex optimization

Stochastic convex optimization

$$\begin{aligned} \text{minimize} \quad & f(x) = \mathbb{E}_{\omega\sim P}[F(x;\omega)] \\ \text{subject to} \quad & x \in \mathcal{X} \end{aligned} \tag{Opt-S}$$

Stochastic convex optimization

Stochastic convex optimization

$$\begin{aligned} \text{minimize} \quad & f(x) = \mathbb{E}_{\omega\sim P}[F(x;\omega)] \\ \text{subject to} \quad & x \in \mathcal{X} \end{aligned} \tag{Opt-S}$$

- Important for data science ⇝ ***finite-sum objectives:***

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

- Special case of OCO:

$$\ell_t \leftarrow f \quad \text{for all } t = 1, 2, \dots$$

- Access to ***stochastic gradients***

$$\hat{g}_t \leftarrow \nabla F(x_t;\omega_t) \quad \text{with } \omega_t \text{ drawn i.i.d. from } P$$

Convergence rate of SGD

Theorem

☞ ***Assume:*** $\mathbb{E}[\|\hat{g}_t\|^2] \le M^2$ *and SGD is run for* T *iterations with* $\eta = \operatorname{diam}(\mathcal{X})/(M\sqrt{T})$

✓ ***Then:*** *the ergodic average* $\bar{x}_T = (1/T)\sum_{t=1}^{T} x_t$ *of SGD enjoys the rate*

$$\mathbb{E}[f(\bar{x}_T) - \min f] \le \frac{M \operatorname{diam}(\mathcal{X})}{\sqrt{T}}$$

Convergence rate of SGD

Theorem

☞ **Assume:** $\mathbb{E}[\|\hat{g}_t\|^2] \leq M^2$ *and SGD is run for T iterations with* $\eta = \mathrm{diam}(\mathcal{X})/(M\sqrt{T})$

✓ **Then:** *the ergodic average* $\bar{x}_T = (1/T)\sum_{t=1}^{T} x_t$ *of SGD enjoys the rate*

$$\mathbb{E}[f(\bar{x}_T) - \min f] \leq \frac{M \,\mathrm{diam}(\mathcal{X})}{\sqrt{T}}$$

Proof.

□

References I

[1] Abernethy, J., Bartlett, P. L., Rakhlin, A., and Tewari, A. Optimal strategies and minimax lower bounds for online convex games. In *COLT '08: Proceedings of the 21st Annual Conference on Learning Theory*, 2008.

[2] Abernethy, J., Lee, C., and Tewari, A. Fighting bandits with a new kind of smoothness. In *NIPS '15: Proceedings of the 29th International Conference on Neural Information Processing Systems*, 2015.

[3] Arora, S., Hazan, E., and Kale, S. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1): 121-164, 2012.

[4] Audibert, J.-Y. and Bubeck, S. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 11: 2635-2686, 2010.

[5] 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.

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

[7] Blackwell, D. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1-8, 1956.

[8] Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1-122, 2012.

[9] Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games.* Cambridge University Press, 2006.

[10] Fudenberg, D. and Levine, D. K. *The Theory of Learning in Games*, volume 2 of *Economic learning and social evolution.* MIT Press, Cambridge, MA, 1998.

References II

[11] 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*, 2021.

[12] Hall, P. and Heyde, C. C. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York, 1980.

[13] Hannan, J. Approximation to Bayes risk in repeated play. In Dresher, M., Tucker, A. W., and Wolfe, P. (eds.), *Contributions to the Theory of Games, Volume III*, volume 39 of *Annals of Mathematics Studies*, pp. 97–139. Princeton University Press, Princeton, NJ, 1957.

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

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

[16] Kakutani, S. A generalization of Brouwer's fixed point theorem. *Duke Mathematical Journal*, 8(3):457–459, September 1941.

[17] Koutsoupias, E. and Papadimitriou, C. H. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pp. 404–413, 1999.

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

[19] Monderer, D. and Shapley, L. S. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996.

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

[21] Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259, 2009.

[22] Ritzberger, K. The theory of normal form games from the differentiable viewpoint. *International Journal of Game Theory*, 23:207–236, September 1994.

References III

[23] Rosenthal, R. W. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65-67, 1973.

[24] Samuelson, L. and Zhang, J. Evolutionary stability in asymmetric games. *Journal of Economic Theory*, 57:363-391, 1992.

[25] Sandholm, W. H. *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, MA, 2010.

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

[27] 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.

[28] Sorin, S. Exponential weight algorithm in continuous time. *Mathematical Programming*, 116(1):513-528, 2009.

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

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

[31] Wilson, R. Computing equilibria of n-person games. *SIAM Journal on Applied Mathematics*, 21:80-87, 1971.

[32] Xiao, L. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11: 2543-2596, October 2010.