THE LIMITS OF REGULARIZED LEARNING IN GAMES

Panayotis Mertikopoulos

French National Center for Scientific Research (CNRS)

Criteo AI Lab

⟨ Learning in the Presence of Strategic Behavior | UC Berkeley | March 28, 2022 ⟩

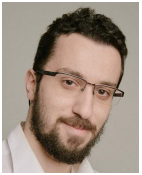
About


A. Giannou


C. Papadimitriou


G. Piliouras


W. H. Sandholm


M. Vlatakis


Z. Zhou

- Giannou, Vlatakis-Gkaragkounis & M, ***Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information,*** COLT 2021
- Giannou, Vlatakis-Gkaragkounis & M, ***The convergence rate of regularized learning in games: From bandits and uncertainty to optimism and beyond,*** NeurIPS 2021
- M, Papadimitriou & Piliouras, ***Cycles in adversarial regularized learning,*** SODA 2018
- M & Sandholm, ***Learning in games via reinforcement and regularization,*** Mathematics of Operations Research, vol. 41, no. 4, pp. 1297-1324, Nov. 2016.
- M & Zhou, ***Learning in games with continuous action sets and unknown payoff functions,*** Mathematical Programming, vol. 173, pp. 465-507, Jan. 2019

Outline

1 Background & Prelims

2 Learning in continuous time

3 Learning in discrete time

Game of roads

A beautiful morning commute in the Bay Area

Online learning

A generic **online decision process:**

---

repeat

At each epoch t

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

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

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

until end

Defining elements

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

Online learning

A generic **online decision process:**

---

repeat

At each epoch t

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

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

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

until end

Defining elements

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

Game-theoretic learning

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

Game-theoretic learning

- **Multiple agents**, individual objectives

[Select a route from home to work]

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

[Encounter other commuters on the road]

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

[Update road choice tomorrow]

Game-theoretic learning

- **Multiple agents**, individual objectives

  [Select a route from home to work]

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

  [Encounter other commuters on the road]

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

  [Update road choice tomorrow]

Does learning lead to stable / rational outcomes?

Finite games in normal form

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

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

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

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

$$v_i(a) \equiv v_i(a_1, \dots, a_N) := (u_i(a_i'; a_{-i}))_{a_i' \in \mathcal{A}_i}$$

[☞ vector of "what-if" / counterfactual rewards]

[☞ $v_i(a)$ *only* depends on a_{-i}]

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

Mixed extensions

Mixed extension of a finite game:

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

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

[x_{ia_i} = prob. that player i plays $a_i \in \mathcal{A}_i$]

- ***Mixed payoff*** of player i

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

[expected payoff of player i under x]

- ***Mixed payoff vector*** of player i:

$$v_i(x) \equiv v_i(x_1, \ldots, x_N) \coloneqq (u_i(a_i; x_{-i}))_{a_i \in \mathcal{A}_i}$$

[☞ vector of *mixed* counterfactual rewards]

[☞ $v_i(x)$ *only* depends on x_{-i}]

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

Nash equilibrium

Nash equilibrium [Nash, 1950, 1951]

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

Nash equilibrium

Nash equilibrium [Nash, 1950, 1951]

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

- **Mixed strategy formulation:**

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

- **Pure strategy formulation:**

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

Nash equilibrium

Nash equilibrium [Nash, 1950, 1951]

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

- **Mixed strategy formulation:**

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

- **Pure strategy formulation:**

$$v_{ia_i^*}(x^*) \geq v_{ia_i}(x^*) \qquad \text{for all } a_i^* \in \operatorname{supp}(x_i^*),\ a_i \in \mathcal{A}_i,\ i \in \mathcal{N}$$

Nash equilibrium

Nash equilibrium [Nash, 1950, 1951]

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

- **Mixed strategy formulation:**

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

- **Pure strategy formulation:**

$$v_{ia_i^*}(x^*) \geq v_{ia_i}(x^*) \qquad \text{for all } a_i^* \in \operatorname{supp}(x_i^*),\ a_i \in \mathcal{A}_i,\ i \in \mathcal{N}$$

- **Pure equilibrium:** $\operatorname{supp}(x^*) = \text{singleton}$ [$x^* = a^* \in \mathcal{A}$]
- **Strict equilibrium:** "$>$" instead of "$\geq$" where appropriate [unique best response; necessarily pure]

Nash equilibrium

Nash equilibrium [Nash, 1950, 1951]

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

- **Mixed strategy formulation:**

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

- **Pure strategy formulation:**

$$v_{ia_i^*}(x^*) \geq v_{ia_i}(x^*) \qquad \text{for all } a_i^* \in \operatorname{supp}(x_i^*),\ a_i \in \mathcal{A}_i,\ i \in \mathcal{N}$$

- **Pure equilibrium:** $\operatorname{supp}(x^*) = \text{singleton}$ [$x^* = a^* \in \mathcal{A}$]
- **Strict equilibrium:** "$>$" instead of "$\geq$" where appropriate [unique best response; necessarily pure]

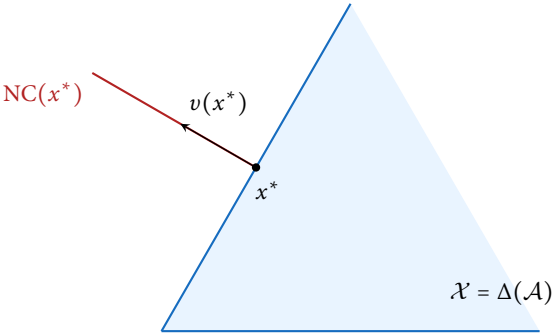
Variational formulation [Stampacchia, 1964]

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

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

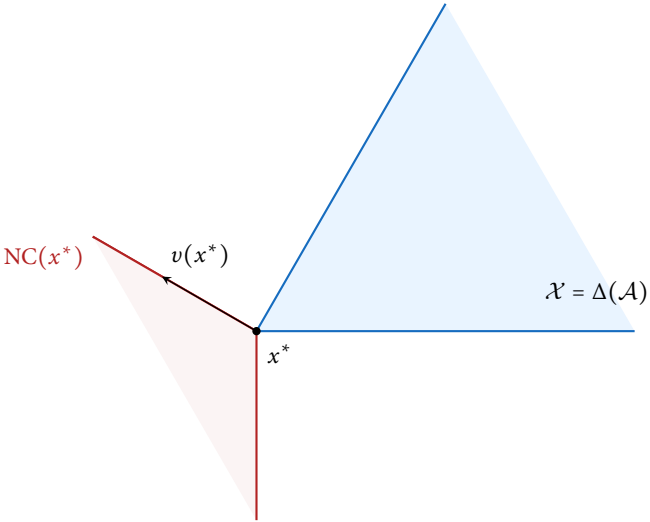
Geometric interpretation

Figure: Different Nash equilibrium configurations: ***mixed***



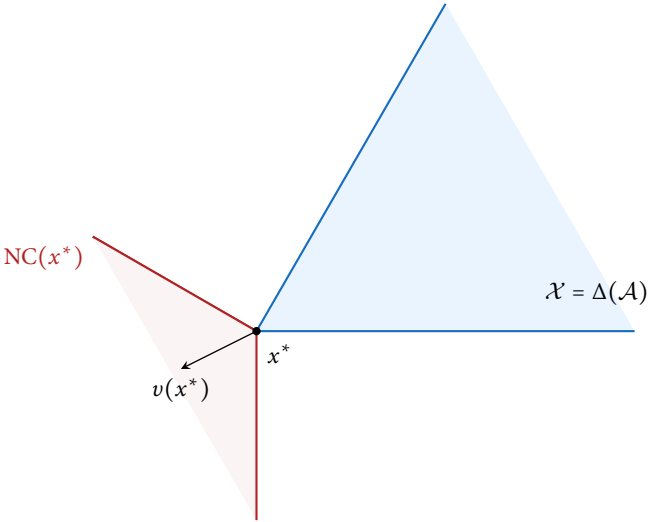
Geometric interpretation

Figure: Different Nash equilibrium configurations: mixed vs. ***pure***



Geometric interpretation

Figure: Different Nash equilibrium configurations: mixed vs. pure vs. ***strict***



Rationality analysis

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

- Are they dynamically ***stable?***
- Are they ***attracting?***
- Do the dynamics converge globally? Locally?
- What other behaviors can we observe?
- ...

Rationality analysis

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

- Are they dynamically ***stable?***
- Are they ***attracting?***
- Do the dynamics converge globally? Locally?
- What other behaviors can we observe?
- ...

Theorem (Hart & Mas-Colell, 2000, informal)

There exist no uncoupled dynamics which guarantee convergence to Nash equilibrium in all games.[a]

[a] Uncoupled $\rightsquigarrow$ the adjustment of a player's strategy does not depend on the payoff functions of the other players.

Outline

1 Background & Prelims

2 Learning in continuous time

3 Learning in discrete time

Learning in 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*** $x_i(t) \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # mixed extension

Encounter **payoff vector** $v_i(x(t))$

Receive reward $u_t(x_t) = \langle v_t, x_t \rangle$ # mixed payoffs

until end

Defining elements

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

Learning with exponential weights

- Agents record cumulative payoff of each strategy

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

$\Longrightarrow$ ***propensity*** of choosing a strategy [Littlestone & Warmuth, 1994; Auer et al., 1995; Sorin, 2009]

Learning with exponential weights

- Agents record cumulative payoff of each strategy

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

$\Longrightarrow$ ***propensity*** of choosing a strategy [Littlestone & Warmuth, 1994; Auer et al., 1995; Sorin, 2009]

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

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

Learning with exponential weights

- Agents record cumulative payoff of each strategy

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

$\Longrightarrow$ ***propensity*** of choosing a strategy [Littlestone & Warmuth, 1994; Auer et al., 1995; Sorin, 2009]

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

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

Learning with exponential weights

- Agents record cumulative payoff of each strategy

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

$\Longrightarrow$ ***propensity*** of choosing a strategy [Littlestone & Warmuth, 1994; Auer et al., 1995; Sorin, 2009]

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

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

- Evolution of mixed strategies

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

Learning with exponential weights

- Agents record cumulative payoff of each strategy

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

$\Longrightarrow$ ***propensity*** of choosing a strategy [Littlestone & Warmuth, 1994; Auer et al., 1995; Sorin, 2009]

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

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

- Evolution of mixed strategies

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

Replicator dynamics (Taylor & Jonker, 1978; Rustichini, 1999)

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

Regularized learning

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

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

Regularized learning

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

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

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

Regularized learning

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

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

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

- **Regularized best responses**

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

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

Regularized learning

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

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

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

- **Regularized best responses**

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

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

Follow the regularized leader (FTRL)

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

The projection dynamics

Example: Quadratic (Euclidean) regularization

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

The projection dynamics

Example: Quadratic (Euclidean) regularization

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

Choice map $\rightsquigarrow$ closest point projection:

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

The projection dynamics

Example: Quadratic (Euclidean) regularization

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

Choice map $\rightsquigarrow$ closest point projection:

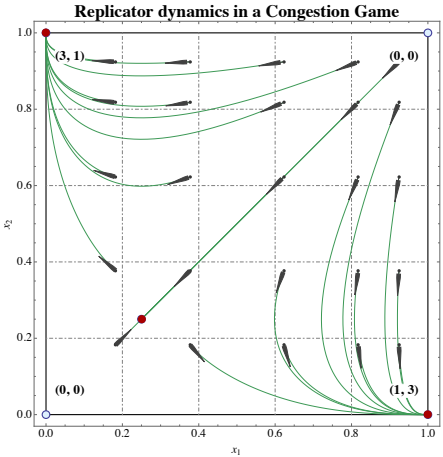
$$\Pi(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - (1/2)\|x\|_2^2\} = \arg\min_{x \in \mathcal{X}} \|y - x\|$$

Projection dynamics [M & Sandholm, 2016]

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

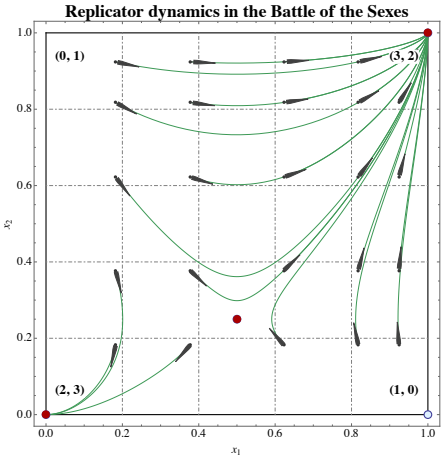
Phase portraits

What do the dynamics look like?



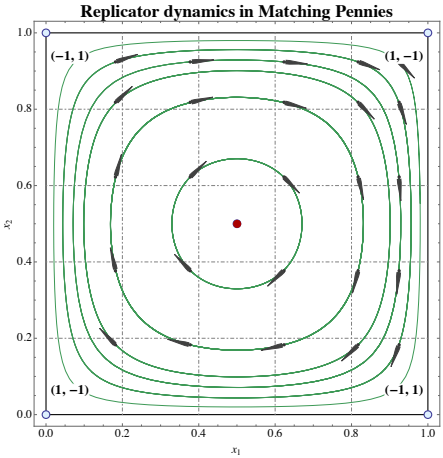
Phase portraits

What do the dynamics look like?



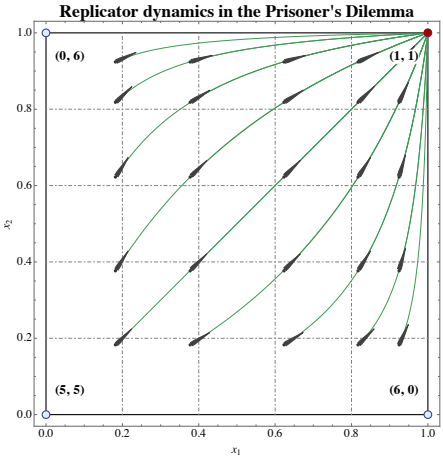
Phase portraits

What do the dynamics look like?



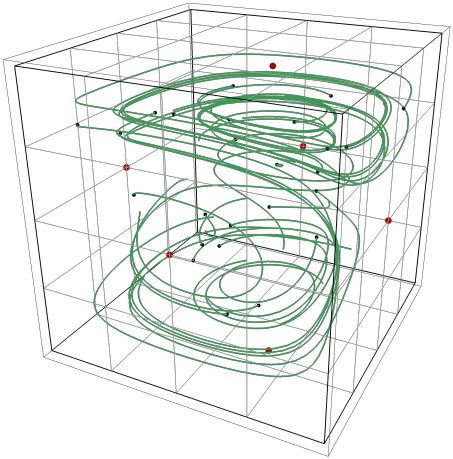
Phase portraits

What do the dynamics look like?



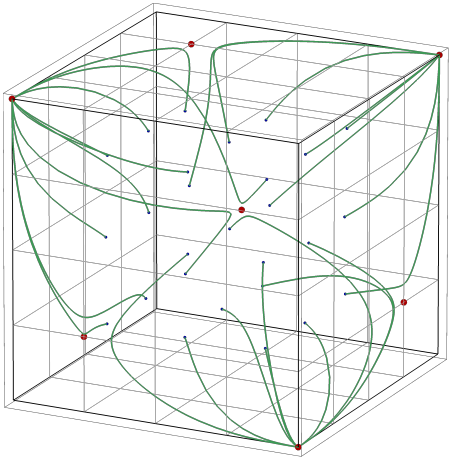
Phase portraits

What do the dynamics look like?



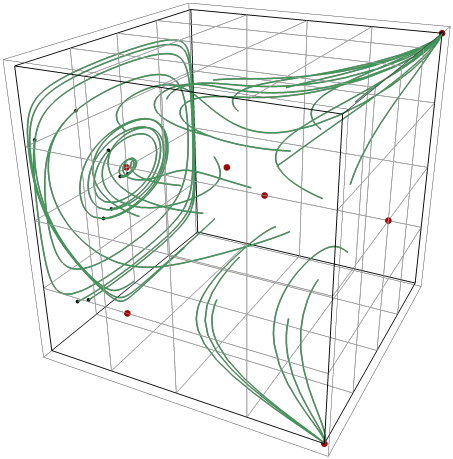
Phase portraits

What do the dynamics look like?



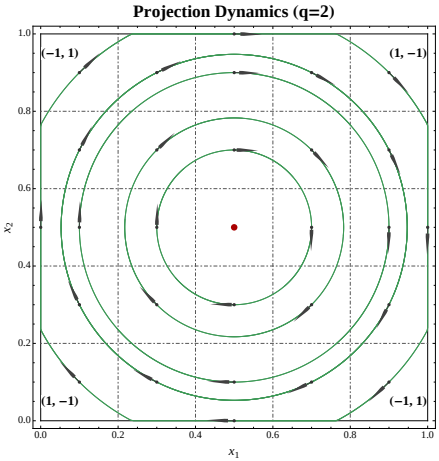
Phase portraits

What do the dynamics look like?



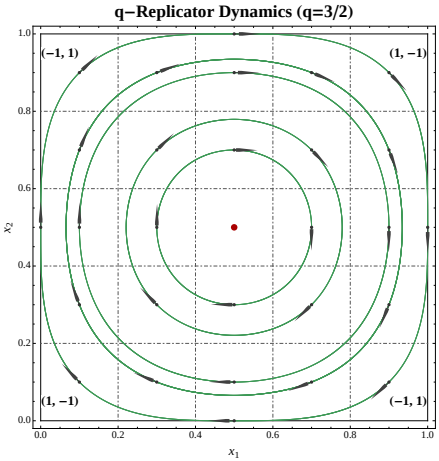
Portraits and examples

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



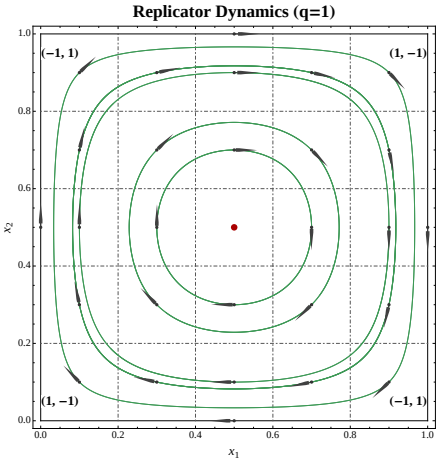
Portraits and examples

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



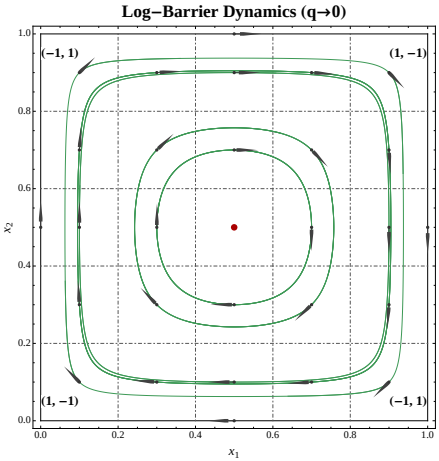
Portraits and examples

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

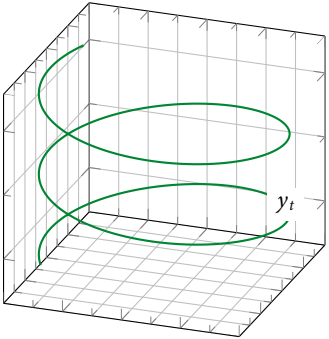


Portraits and examples

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

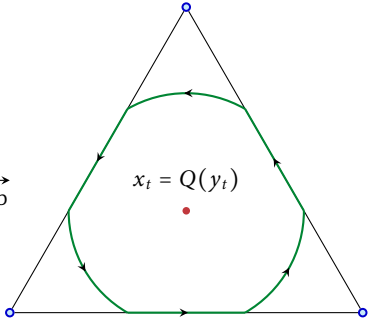


In and out of the boundary



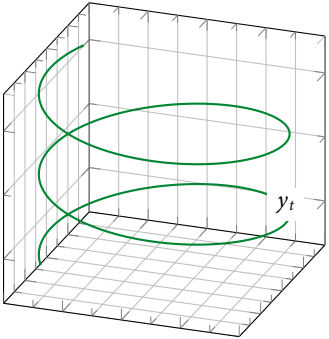
Payoff space (dual)

$$\xrightarrow[\text{choice map}]{Q}$$



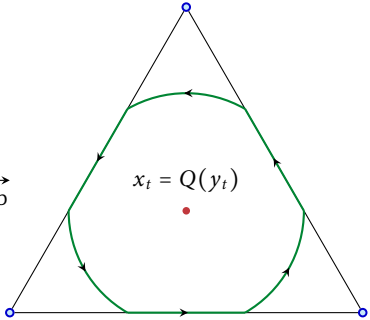
Strategy space (primal)

In and out of the boundary



Payoff space (dual)

$$\xrightarrow[\text{choice map}]{Q}$$



Strategy space (primal)

Key difference with replicator: faces no longer forward invariant

Stationarity of equilibria

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

- Supported strategies have equal payoffs:

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

- Mean payoff equal to equilibrium payoff:

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

- Replicator field vanishes at Nash equilibria:

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

Proposition (Stationarity of Nash equilibria)

Let $x(t) = Q(y(t))$ be a solution orbit of (C-FTRL). Then:

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

Stationarity of equilibria

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

- Supported strategies have equal payoffs:

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

- Mean payoff equal to equilibrium payoff:

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

- Replicator field vanishes at Nash equilibria:

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

Proposition (Stationarity of Nash equilibria)

Let $x(t) = Q(y(t))$ be a solution orbit of (C-FTRL). Then:

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

✗ The converse does not hold!

[See previous portraits]

Stability

Are all stationary points created equal?

Definition (Notions of 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]
- x^* is ***attracting*** if $\lim_{t\to\infty} x(t) = x^*$ whenever $x(0)$ is close enough to x^*
[Trajectories that start close to x^* eventually converge to x^*]
- x^* is ***asymptotically stable*** if it is stable and attracting

A "folk theorem" for learning

Are all equilibria created equal?

Theorem

Let Γ be a finite game and let $x(t) = Q(y(t))$ be a solution orbit of (C-FTRL). Then:

1. x^* *is a Nash equilibrium* $\implies$ x^* *is stationary* [M & Sandholm, 2016]
2. $\lim_{t\to\infty} x(t) = x^*$ $\implies$ x^* *is a Nash equilibrium* [M & Sandholm, 2016]
3. x^* *is stable* $\implies$ x^* *is a Nash equilibrium* [M & Sandholm, 2016]
4. x^* *is a strict Nash equilibrium* $\iff$ x^* *is asymptotically stable* [$\implies$ M & Sandholm, 2016; $\impliedby$ Flokas et al., 2020]

Some remarks:

- ***Only strict equilibria can be stable*** [☞ Vidya's talk]
- For replicator dynamics $\rightsquigarrow$ folk theorem of evolutionary game theory [Hofbauer & Sigmund, 2003]
- Definition of stability/stationarity requires some thought if h non-steep [duality of (C-FTRL)]

Non-convergence in zero-sum games

The min-max case is quite different (and special):

Non-convergence in zero-sum games

The min-max case is quite different (and special):

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

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

Non-convergence in zero-sum games

The min-max case is quite different (and special):

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

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

Theorem (Hofbauer et al., 2009)

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

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

Poincaré recurrence in zero-sum games

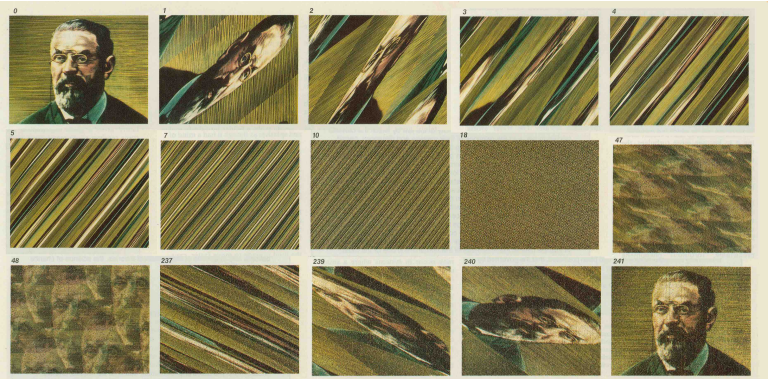
Definition (Poincaré)

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

Poincaré recurrence in zero-sum games

Definition (Poincaré)

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



Poincaré recurrence in zero-sum games

Proposition

(C-FTRL) *is volume-preserving under the Hessian Riemannian metric*

$$g_{aa'}(x) = \frac{\partial^2 h}{\partial x_a \partial x_{a'}}$$

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

Poincaré recurrence in zero-sum games

Proposition

(C-FTRL) *is volume-preserving under the Hessian Riemannian metric*

$$g_{aa'}(x) = \frac{\partial^2 h}{\partial x_a \partial x_{a'}}$$

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

- but the metric becomes **singular** at the boundary of $\mathcal{X}$ if h is steep ✗
- ...and the dynamics may collide with the boundary of $\mathcal{X}$ in finite time otherwise ✗

Poincaré recurrence in zero-sum games

Proposition

(C-FTRL) *is volume-preserving under the Hessian Riemannian metric*

$$g_{aa'}(x) = \frac{\partial^2 h}{\partial x_a \partial x_{a'}}$$

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

- but the metric becomes **singular** at the boundary of $\mathcal{X}$ if h is steep ✗
- ...and the dynamics may collide with the boundary of $\mathcal{X}$ in finite time otherwise ✗

Theorem (M et al., 2018)

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

Outline

1 Background & Prelims

2 Learning in continuous time

3 Learning in discrete time

The model

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

repeat

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

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

Choose ***action*** $\hat{a}_{i,n} \sim X_{i,n}$ — # random action selection

Observe **mixed payoff vector** $v_i(X_n)$ — # feedback phase

until end

Defining elements

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

The model

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

repeat

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

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

Choose ***action*** $\hat{a}_{i,n} \sim X_{i,n}$ — # random action selection

Observe pure payoff vector $v_i(\hat{a}_n)$ — # feedback phase

until end

Defining elements

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

The model

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

repeat

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

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

Choose ***action*** $\hat{a}_{i,n} \sim X_{i,n}$ — # random action selection

Observe **realized payoff** $u_i(\hat{a}_n)$ — # feedback phase

until end

Defining elements

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

The feedback process

Different types of feedback (from best to worst):

- ***Mixed payoff vectors***: $v_i(X_n)$ [vector feedback]
- ***Pure payoff vectors***: $v_i(\hat{a}_n)$ [vector feedback]
- ***Bandit / Payoff-based:*** $u_{i,n}(\hat{a}_n)$ [scalar feedback]

The feedback process

Different types of feedback (from best to worst):

- ***Mixed payoff vectors***: $v_i(X_n)$ [vector feedback]
- ***Pure payoff vectors***: $v_i(\hat{a}_n)$ [vector feedback]
- ***Bandit / Payoff-based:*** $u_{i,n}(\hat{a}_n)$ [scalar feedback]

Features:

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

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

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

From payoffs to payoff vectors

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

From payoffs to payoff vectors

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

Definition (Importance weighted estimation)

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

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

From payoffs to payoff vectors

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

Definition (Importance weighted estimation)

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

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

Statistical properties of (IWE)

- ***Unbiased:***

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

- ***Second moment:***

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

The oracle model

Definition (Black-box oracle)

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

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

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

The oracle model

Definition (Black-box oracle)

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

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

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

Examples

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

Exponential weights redux

Algorithm Exponential weights in discrete time (ExpWeight)

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

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

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

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

play $\hat{a}_{i,n} \sim X_{i,n}$ # choose action

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

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

end for

Basic idea:

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

Model 1: ExpWeight with mixed payoff vector observations

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

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

Oracle features:

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

Model 2: ExpWeight with pure payoff vector observations

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

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

Oracle features:

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

☞ this algorithm is known as as **Hedge** [Auer et al., 1995, 2002b,a]

Model 3: ExpWeight with bandit feedback

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

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

Oracle features:

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

☞ this algorithm is known as as **EXP3** [Auer et al., 1995, 2002b,a]

Model 4: ExpWeight with bandit feedback

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

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

Oracle features:

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

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

Model 5: Optimistic ExpWeight / Multiplicative Weights

If players are ***optimistic***:[1] [DP19]

$$\hat{v}_{i,n} = v_i(X_{i,n+1/2}; X_{-i,n+1/2})$$

Oracle features:

- ***Deterministic:*** no randomness
- ***Bias:*** $B_n = v(X_{n+1/2}) - v(X_n) = \mathcal{O}(\gamma_n)$
- ***Variance:*** $\sigma_n = 0$
- ***Second moment:*** $M_n = \mathcal{O}(1)$

[1] Feedback obtained via the sequence

$$Y_{n+1/2} = Y_n + \gamma_n v_n(X_{n-1/2}) \qquad X_{i,n+1/2} \propto \exp(Y_{i,n+1/2}) \qquad Y_{n+1} = Y_n + \gamma_n v(X_{n+1/2})$$

Model 6: Clairvoyant ExpWeight

If players are ***clairvoyant***: [PSS21]

$$\hat{v}_{i,n} = v_i\big(X_{i,n+1}; X_{-i,n+1}\big)$$

Oracle features:

- ***Deterministic:*** no randomness
- ***Bias:*** $B_n = v(X_{n+1}) - v(X_n) = \mathcal{O}(\gamma_n)$
- ***Variance:*** $\sigma_n = 0$
- ***Second moment:*** $M_n = \mathcal{O}(1)$

Follow the generalized leader

Follow the generalized leader

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

[Shalev-Shwartz & Singer, 2006; Nesterov, 2009]

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

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

Examples

Example (Ridge regularization)

- **Regularizer:**

$$h(x) = \frac{1}{2}\|x\|_2^2$$

- **Algorithm:**

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

Examples

Example (Ridge regularization)

- **Regularizer:**

$$h(x) = \frac{1}{2}\|x\|_2^2$$

- **Algorithm:**

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

Example (Entropic regularization)

- **Regularizer:**

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

- **Algorithm:**

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

Visualization

What does the sequence of play look like?

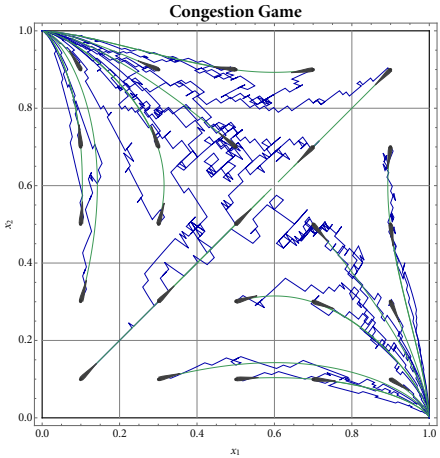


Figure: Optimistic ExpWeight with constant step-size

Visualization

What does the sequence of play look like?

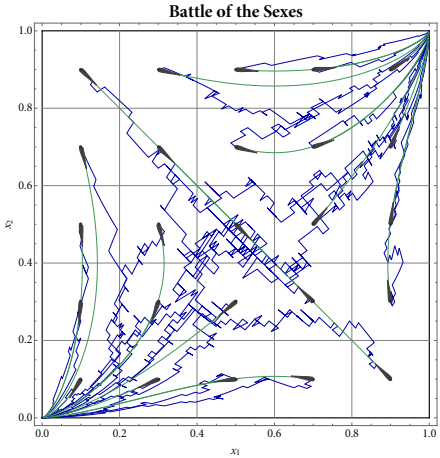


Figure: Optimistic ExpWeight with constant step-size

Visualization

What does the sequence of play look like?

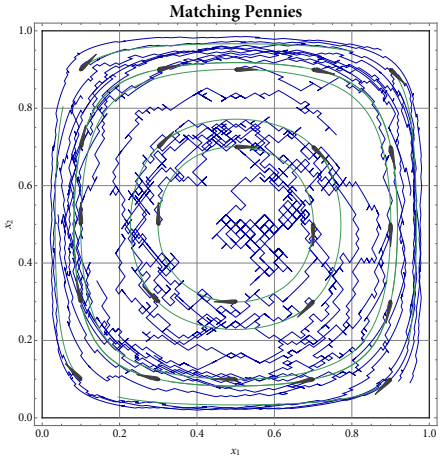


Figure: Optimistic ExpWeight with constant step-size

Visualization

What does the sequence of play look like?

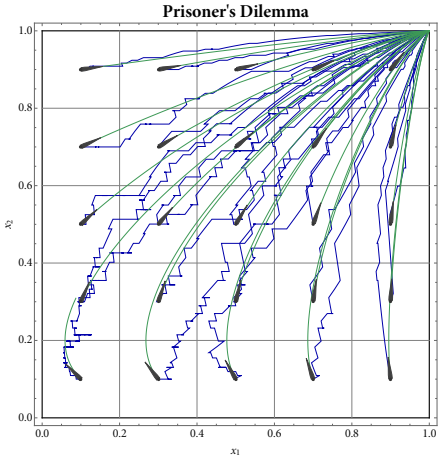


Figure: Optimistic ExpWeight with constant step-size

Notions of stability

Definition (Stochastic stability)

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

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

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

Notions of stability

Definition (Stochastic stability)

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

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

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

Definition (Stochastic asymptotic stability)

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

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

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

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

The long-run behavior of regularized learning

Theorem

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

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

(A2) $b_n \to 0$

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

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

[a] "Only if" direction requires generic noise at equilibrium: $\mathbb{P}(|\hat{v}_{a,n} - \hat{v}_{a',n}| \leq c \mid \mathcal{F}_n) > 0$; satisfied by all stochastic models discussed.

The long-run behavior of regularized learning

Theorem

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

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

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

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

[a] "Only if" direction requires generic noise at equilibrium: $\mathbb{P}(|\hat{v}_{a,n} - \hat{v}_{a',n}| \leq c \mid \mathcal{F}_n) > 0$; satisfied by all stochastic models discussed.

The long-run behavior of regularized learning

Theorem

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

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

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

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

☞ ***Then:*** *the sequence* X_n *generated by* (FTGL) *enjoys the following properties*

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

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

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

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

[a] "Only if" direction requires generic noise at equilibrium: $\mathbb{P}(|\hat{v}_{a,n} - \hat{v}_{a',n}| \leq c \mid \mathcal{F}_n) > 0$; satisfied by all stochastic models discussed.

Rate of convergence

Theorem (Giannou et al., 2021b)

- ***Assume:*** *all players run (FTGL) with step-size γ_n and oracle parameters b_n (bias) and U_n (noise) as before*
- ***Then:*** *if x^* is a strict Nash equilibrium and X_n converges to x^*, we have*

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

where

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

Overview

I. Learning in continuous time

- ✓ Nash equilibrium $\Longrightarrow$ stationarity
- ✓ Lyapunov stability $\Longrightarrow$ equilibrium
- ✓ Asymptotic stability $\Longleftrightarrow$ strict equilibrium
- ✓ Zero-sum games $\Longrightarrow$ Poincaré recurrence

II. Learning in discrete time

- ✗ Depends on feedback, step-size, ... [stochastic ≠ deterministic]
- ✗ Nash equilibrium $\not\Longrightarrow$ stationarity
- ✓ Lyapunov stability $\Longrightarrow$ equilibrium
- ✓ Asymptotic stability $\Longleftrightarrow$ strict equilibrium [mixed equilibria are ***weak***]
- ✗ Zero-sum games $\not\Longrightarrow$ Poincaré recurrence [convergence to different distance]

Open issues

- What if information enters the algorithm with the same weight? [cf. Vidya's work]
- Adaptive step-size / learning rate? [challenging analysis]
- Robustness to delays / corruptions / ...
- Non-singleton attractors? Other limit behaviors? [☞ hic sunt leones]
- Learning in continuous games?

References I

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

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

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

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

Daskalakis, C. and Panageas, I. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *ITCS '19: Proceedings of the 10th Conference on Innovations in Theoretical Computer Science*, 2019.

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

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

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

Hart, S. and Mas-Colell, A. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127-1150, September 2000.

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

References II

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

Hsieh, Y.-P., Mertikopoulos, P., and Cevher, V. The limits of min-max optimization algorithms: Convergence to spurious non-critical sets. In *ICML '21: Proceedings of the 38th International Conference on Machine Learning*, 2021.

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

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

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

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

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

Mertikopoulos, P., Hallak, N., Kavis, A., and Cevher, V. On the almost sure convergence of stochastic gradient descent in non-convex problems. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

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

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

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

References III

Piliouras, G., Sim, R., and Skoulakis, S. Optimal no-regret learning in general games: Bounded regret with unbounded step-sizes via clairvoyant mwu. https://arxiv.org/abs/2111.14737, 2021.

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

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

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

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

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

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