LEARNING IN HARMONIC GAMES

Panayotis Mertikopoulos

French National Center for Scientific Research (CNRS)

⟨ Workshop on Learning in Games | Toulouse | July 3, 2024⟩

About

- Legacci, M, & Pradelski, *A geometric decomposition of finite games: Convergence vs. recurrence under exponential weights,* ICML 2024
- Legacci, M, Papadimitriou, Piliouras, & Pradelski, *No-regret learning in harmonic games: Extrapolation in the presence of conflicting interests,* preprint, 2024

D. Legacci

B. Pradelski

About

- Legacci, M, & Pradelski, *A geometric decomposition of finite games: Convergence vs. recurrence under exponential weights,* ICML 2024
- Legacci, M, Papadimitriou, Piliouras, & Pradelski, *No-regret learning in harmonic games: Extrapolation in the presence of conflicting interests,* preprint, 2024


D. Legacci


C. Papadimitriou


G. Piliouras


B. Pradelski

Outline

Finite games

Finite games

A ***finite game*** $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ consists of

- A finite set of ***players*** $\mathcal{N} = \{1, \dots, N\}$
- A finite set of ***actions*** (or ***pure strategies***) $\mathcal{A}_i = \{1, \dots, A_i\}$ per player $i \in \mathcal{N}$
- An ensemble of ***payoff functions*** $u_i \colon \mathcal{A} \equiv \prod_i \mathcal{A}_i \to \mathbb{R}$, $i \in \mathcal{N}$

Notation

- ***Action profile:***

$$\alpha \equiv (\alpha_1, \dots, \alpha_N) \in \mathcal{A} := \prod_{i \in \mathcal{N}} \mathcal{A}_i$$

- ***Pure*** - or ***realized*** - ***payoff*** of player i

$$u_i(\alpha) \equiv u_i(\alpha_i; \alpha_{-i}) \coloneqq u_i(\alpha_1, \dots, \alpha_N)$$

- ***Pure*** - or ***realized*** - ***payoff vector*** of player i

$$v_i(\alpha) \equiv v_i(\alpha_1, \dots, \alpha_N) \coloneqq (u_i(\beta_i; \alpha_{-i}))_{\beta_i \in \mathcal{A}_i}$$

vector of "what-if" / counterfactual rewards

Mixed extensions

Players can ***randomize***

- ***Mixed strategy*** of player i

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

$x_{i\alpha_i}$ = prob. that player i plays $\alpha_i \in \mathcal{A}_i$

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

$$u_i(x) = \mathbb{E}_{\alpha \sim x}[u_i(\alpha)] = \sum_{\alpha_1 \in \mathcal{A}_1} \dots \sum_{\alpha_N \in \mathcal{A}_N} x_{1,\alpha_1} \cdots x_{N,\alpha_N} \, u_i(\alpha_1, \dots, \alpha_N)$$

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

$$v_i(x) \coloneqq (u_i(\alpha_i; x_{-i}))_{\alpha_i \in \mathcal{A}_i}$$

vector of *expected* rewards
$v_i(x)$ *only* depends on x_{-i}

- ***Mixed extension of*** Γ: $\bar{\Gamma} \equiv \Delta(\Gamma)$

Nash equilibrium

Nash equilibrium [Nash, 1950]

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

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

Nash equilibrium

Nash equilibrium [Nash, 1950]

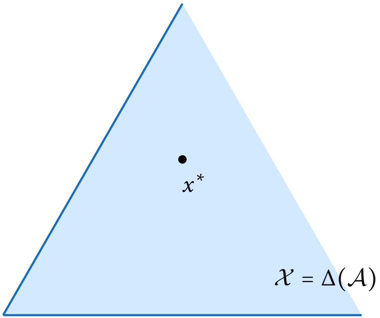
"No player has an incentive to deviate from their chosen strategy if other players don't"

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

- **Variational characterization:** $\langle v(x^*), x - x^* \rangle \leq 0$ for all $x \in \mathcal{X}$ # Stampacchia variational inequality
- **Pure equilibrium:** $\operatorname{supp}(x^*) = \text{singleton}$ # pure strategy profile
- **Strict equilibrium:** "$>$" instead of "$\geq$" # unique best response; necessarily pure

Equilibrium configurations

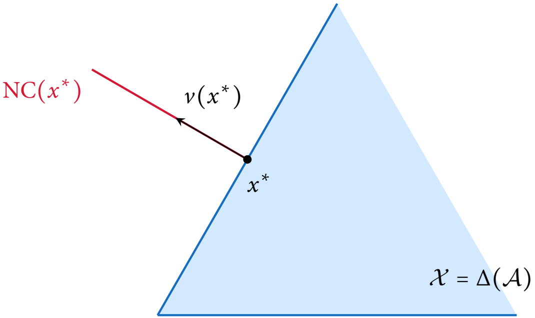
Figure: Different equilibrium configurations: ***fully mixed***



$\mathrm{NC}(x^*)$ = normal cone at x^* (outward normal directions)

Equilibrium configurations

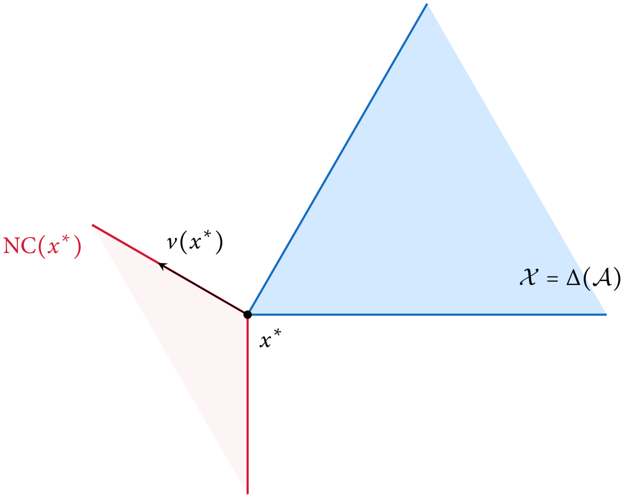
Figure: Different equilibrium configurations: fully mixed vs. ***mixed***



$\mathrm{NC}(x^*)$ = normal cone at x^* (outward normal directions)

Equilibrium configurations

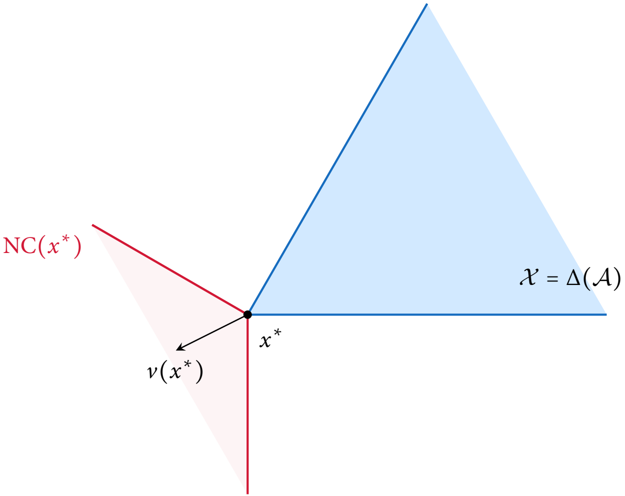
Figure: Different equilibrium configurations: fully mixed vs. mixed vs. ***pure***



NC(x^*) = normal cone at x^* (outward normal directions)

Equilibrium configurations

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



$\mathrm{NC}(x^*)$ = normal cone at x^* (outward normal directions)

Potential games

Potential games [Monderer & Shapley, 1996]

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

$$u_i(\alpha_i; \alpha_{-i}) - u_i(\beta_i; \alpha_{-i}) = \Phi(\alpha_i; \alpha_{-i}) - \Phi(\beta_i; \alpha_{-i}) \quad \text{for all } \alpha, \beta \in \mathcal{A}$$

Potential games

Potential games [Monderer & Shapley, 1996]

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

$$u_i(\alpha_i; \alpha_{-i}) - u_i(\beta_i; \alpha_{-i}) = \Phi(\alpha_i; \alpha_{-i}) - \Phi(\beta_i; \alpha_{-i}) \quad \text{for all } \alpha, \beta \in \mathcal{A}$$

Basic properties

- Player interests **aligned** along a common objective # common interest
- Improvement paths **always** terminate # no best-response cycles
- **Always admit pure equilibria** # generically **strict**

Graphical representation

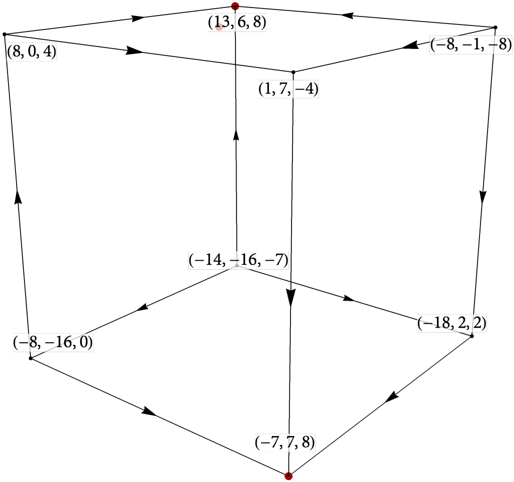


Figure: Response graph of a potential game (Nash in red)

Graphical representation

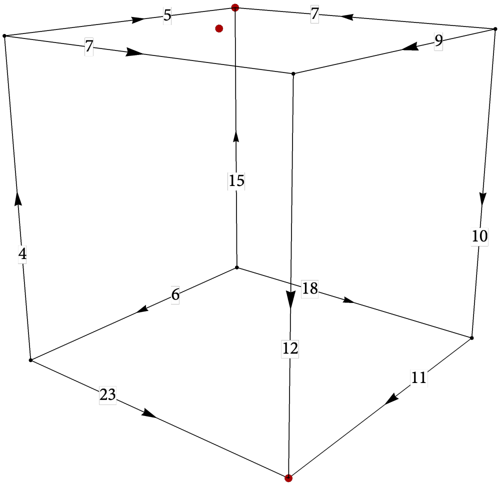


Figure: Response graph of a potential game (Nash in red)

Harmonic games

Harmonic games [Candogan et al., 2011; Abdou et al., 2022]

The game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ is a ***harmonic game*** if it admits a ***harmonic measure*** $\mu\colon \coprod_i \mathcal{A}_i \to (0, \infty)$ such that

$$\sum_{i\in\mathcal{N}} \sum_{\beta\in\mathcal{A}_i} \mu_{i\beta_i} [u_i(\alpha_i; \alpha_{-i}) - u_i(\beta_i; \alpha_{-i})] = 0 \quad \text{for all } \alpha \in \mathcal{A}$$

☞ ***uniform harmonic:*** $\mu_i = \mathrm{unif}_{\mathcal{A}_i}$

☞ ***flow conservation:*** no deviation sources or sinks

☞ terminology: harmonic component of Hodge decomposition

Harmonic games

Harmonic games [Candogan et al., 2011; Abdou et al., 2022]

The game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$ is a ***harmonic game*** if it admits a ***harmonic measure*** $\mu\colon \coprod_i \mathcal{A}_i \to (0, \infty)$ such that

$$\sum_{i \in \mathcal{N}} \sum_{\beta \in \mathcal{A}_i} \mu_{i\beta_i} [u_i(\alpha_i; \alpha_{-i}) - u_i(\beta_i; \alpha_{-i})] = 0 \quad \text{for all } \alpha \in \mathcal{A}$$

☞ ***uniform harmonic:*** $\mu_i = \mathrm{unif}_{\mathcal{A}_i}$

☞ ***flow conservation:*** no deviation sources or sinks

☞ terminology: harmonic component of Hodge decomposition

Basic properties

- Player interests **anti-aligned** # conflicts of interest
- Improvement paths **never** terminate # best-response sinks
- **No pure equilibria** # at least one fully mixed

Graphical representation

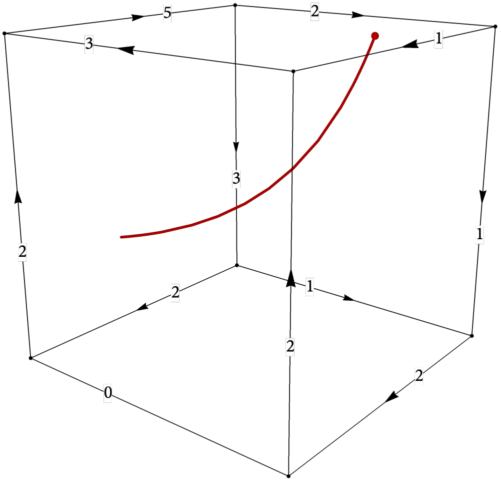


Figure: Response graph of a harmonic game (Nash in red)

Harmonic resonance

Examples of harmonic games

- Matching Pennies, Rock-Paper-Scissors, Dawkins' battle of the sexes, ...
- Two-player zero-sum games with a fully mixed equilibrium — Legacci et al. (2024)
- Cyclic games — Hofbauer & Schlag (2000)
- ...

Harmonic resonance

Examples of harmonic games

- Matching Pennies, Rock-Paper-Scissors, Dawkins' battle of the sexes, ...
- Two-player zero-sum games with a fully mixed equilibrium — Legacci et al. (2024)
- Cyclic games — Hofbauer & Schlag (2000)
- ...

Hodge decomposition of games [Candogan et al., 2011; Abdou et al., 2022]

Any finite game Γ *can be decomposed as*

$$\Gamma = \Gamma_{\text{pot}} + \Gamma_{\text{harm}}$$

where Γ_{pot} *is potential and* Γ_{harm} *is harmonic*

Harmonic resonance

Hodge decomposition of games [Candogan et al., 2011; Abdou et al., 2022]

Any finite game Γ *can be decomposed as*

$$\Gamma = \Gamma_{\mathrm{pot}} + \Gamma_{\mathrm{harm}}$$

where Γ_{pot} *is potential and* Γ_{harm} *is harmonic*

Harmonic resonance

Hodge decomposition of games [Candogan et al., 2011; Abdou et al., 2022]

Any finite game Γ *can be decomposed as*

$$\Gamma = \Gamma_{\text{pot}} + \Gamma_{\text{harm}}$$

where Γ_{pot} *is potential and* Γ_{harm} *is harmonic*

Remarks:

- Decomposition **not unique** # must fix measure / gauge
- Harmonic and potential games are **orthogonal** # given a measure / metric
- ***Harmonic*** $\neq$ ***zero-sum*** ⚠ # zero-sum can be potential

Harmonic resonance

Hodge decomposition of games [Candogan et al., 2011; Abdou et al., 2022]

Any finite game Γ can be decomposed as

$$\Gamma = \Gamma_{\text{pot}} + \Gamma_{\text{harm}}$$

where Γ_{pot} is potential and Γ_{harm} is harmonic

Remarks:

- Decomposition **not unique** — # must fix measure / gauge
- Harmonic and potential games are **orthogonal** — # given a measure / metric
- ***Harmonic $\neq$ zero-sum*** ⚠ — # zero-sum can be potential

Harmonic games $\rightsquigarrow$ strategic complement of potential games

This talk

Are harmonic and potential games complementary from a dynamic viewpoint?

What is the long-run behavior of no-regret learning in harmonic games?

Outline

Multi-agent learning

Sequence of events - generic

for each ***epoch*** and every ***player*** **do**	# continuous / discrete
Choose ***action***	# continuous / finite
Receive ***reward***	# endogenous / exogenous
Get ***feedback*** (maybe)	# full info / oracle / payoff-based
end for	

Defining elements

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

Multi-agent learning

Sequence of events - generic

for each ***epoch*** and every ***player*** **do** — # continuous / discrete

  Choose ***action*** — # continuous / finite

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

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

end for

Defining elements

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

Regret minimization

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

Regret minimization

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

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

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

Worst-case guarantee: at the very least, minimize regret

Regret minimization

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

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

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

Worst-case guarantee: at the very least, minimize regret

Literature:

- Economics — Hannan (1957); Hart & Mas-Colell (2000); Fudenberg & Levine (1998)
- Mathematics — Blackwell (1956); Lai & Robbins (1985); Sorin (2024)
- Computer science — Shalev-Shwartz (2011); Cesa-Bianchi & Lugosi (2006); Lattimore & Szepesvári (2020)

Regret minimization and rationality

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

Does no-regret learning converge to equilibrium?

Regret minimization and rationality

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

Under no-regret learning, **empirical frequencies of play** converge to equilibrium

Regret minimization and rationality

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

Under no-regret learning, **empirical frequencies of play** converge to the game's **Hannan set / set of coarse correlated equilibria**

Regret minimization and rationality

Individual regret

$$\mathrm{Reg}_i(T) = \max_{\alpha_i \in \mathcal{A}_i} \sum_{t=1}^{T} [u_i(\alpha_i; z_{-i,t}) - u_i(z_t)]$$

Under no-regret learning, **empirical frequencies of play** converge to the game's **Hannan set / set of coarse correlated equilibria**

Empirical frequencies of play

$$z_{\alpha,t} = \frac{1}{t} \sum_{\tau=1}^{t} \mathbb{1}\{\alpha_\tau = \alpha\}$$

Coarse correlated equilibrium

A correlated strategy $z \in \Delta(\mathcal{A})$ is ***a coarse correlated equilibrium / Hannan consistent*** if

$$u_i(z) \geq u_i(\beta_i; z_{-i}) \quad \text{for all } \beta_i \in \mathcal{A}_i, i \in \mathcal{N}$$

The VZ game

A ***coordination game***

	A	B
A	$(1,1)$	$(0,0)$
B	$(0,0)$	$(1,1)$

The VZ game

A ***coordination game*** **with two feeble twins**

Viossat & Zapechelnyuk, 2013

	A	A_-	B	B_-
A	$(1,1)$	$(1,2/3)$	$(0,0)$	$(0,-1/3)$
A_-	$(2/3,1)$	$(2/3,2/3)$	$(-1/3,0)$	$(-1/3,-1/3)$
B	$(0,0)$	$(0,-1/3)$	$(1,1)$	$(1,2/3)$
B_-	$(-1/3,0)$	$(-1/3,-1/3)$	$(2/3,1)$	$(2/3,2/3)$

Feeble twins:

- A_- is **strictly $1/3$-dominated** by A
- B_- is **strictly $1/3$-dominated** by B

The VZ game

A ***coordination game*** **with two feeble twins** Viossat & Zapechelnyuk, 2013

	A	A_-	B	B_-
A	$(1,1)$	$(1,2/3)$	$(0,0)$	$(0,-1/3)$
A_-	$(2/3,1)$	$(2/3,2/3)$	$(-1/3,0)$	$(-1/3,-1/3)$
B	$(0,0)$	$(0,-1/3)$	$(1,1)$	$(1,2/3)$
B_-	$(-1/3,0)$	$(-1/3,-1/3)$	$(2/3,1)$	$(2/3,2/3)$

Feeble twins:

- A_- is **strictly $1/3$-dominated** by A
- B_- is **strictly $1/3$-dominated** by B

BUT!

- Suppose players play (A_-, A_-) and (B_-, B_-) each with prob. $1/2$
- ***Distribution of play is a CCE:*** $u_i(\alpha_i; z_{-i}) - u_i(z) \leq -1/6$
- ***No regret!*** # in fact, ***negative*** regret

The VZ game

A ***coordination game*** **with two feeble twins** — Viossat & Zapechelnyuk, 2013

	A	A_-	B	B_-
A	$(1,1)$	$(1,2/3)$	$(0,0)$	$(0,-1/3)$
A_-	$(2/3,1)$	$(2/3,2/3)$	$(-1/3,0)$	$(-1/3,-1/3)$
B	$(0,0)$	$(0,-1/3)$	$(1,1)$	$(1,2/3)$
B_-	$(-1/3,0)$	$(-1/3,-1/3)$	$(2/3,1)$	$(2/3,2/3)$

Feeble twins:

- A_- is **strictly $1/3$-dominated** by A
- B_- is **strictly $1/3$-dominated** by B

BUT!

- Suppose players play (A_-, A_-) and (B_-, B_-) each with prob. $1/2$
- ***Distribution of play is a CCE:*** $u_i(\alpha_i; z_{-i}) - u_i(z) \leq -1/6$
- ***No regret!*** # in fact, ***negative*** regret

No-regret play may lead to playing dominated strategies for all time!

Outline

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

Get **mixed payoff** $u_i(x(t)) = \langle v_i(x(t)), x_i(t) \rangle$ — # payoff phase

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

until end

Defining elements

- **Time:** $t \geq 0$
- **Players:** many (finite) — # multi-agent learning
- **Actions:** finite
- **Payoffs:** endogenous — # game-theoretic learning
- **Feedback:** mixed payoff vectors

Learning with exponential weights

Exponential reinforcement mechanism:

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

$$y_{i\alpha_i}(t) = \int_0^t v_{i\alpha_i}(x(\tau))\, d\tau$$

- Play an action with probability exponentially proportional to its score

$$x_{i\alpha_i}(t) \propto \exp(y_{i\alpha_i}(t))$$

Exponential weight dynamics [Littlestone & Warmuth, 1994; Auer et al., 1995]

$$\dot{y}_i = v_i(x) \qquad x_i = \Lambda(y_i) \coloneqq \frac{\exp(y_i)}{\|\exp(y_i)\|_1} \tag{EWD}$$

Learning with exponential weights

Exponential reinforcement mechanism:

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

$$y_{i\alpha_i}(t) = \int_0^t v_{i\alpha_i}(x(\tau))\, d\tau$$

- Play an action with probability exponentially proportional to its score

$$x_{i\alpha_i}(t) \propto \exp(y_{i\alpha_i}(t))$$

The replicator dynamics [Taylor & Jonker, 1978]

$$\dot{x}_{i\alpha_i} = x_{i\alpha_i}[u_i(\alpha_i; x_{-i}) - u_i(x)] \tag{RD}$$

What do the dynamics look like?

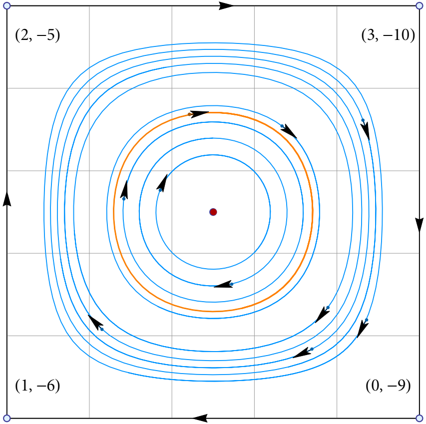


Figure: The replicator dynamics in a 2 × 2 harmonic game (Nash in red)

What do the dynamics look like?

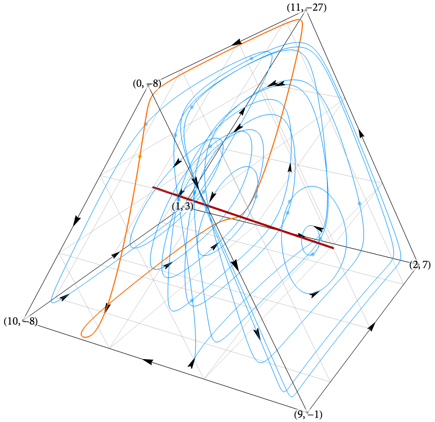


Figure: The replicator dynamics in a 2 × 3 harmonic game (Nash in red)

What do the dynamics look like?

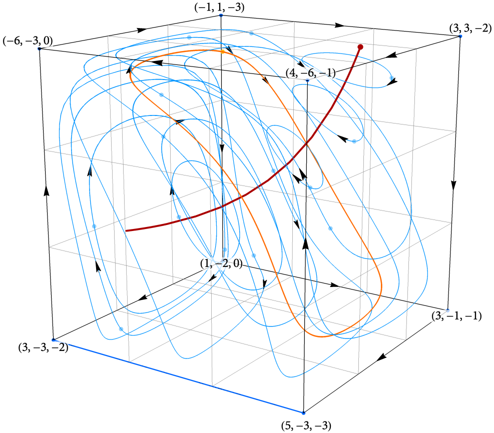


Figure: The replicator dynamics in a 2 × 2 × 2 harmonic game (Nash in red)

What do the dynamics look like?

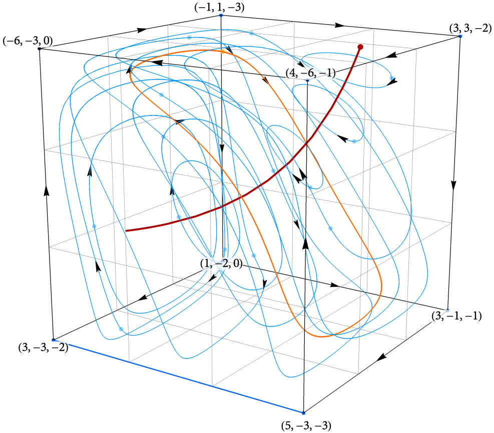


Figure: The replicator dynamics in a $2 \times 2 \times 2$ harmonic game (Nash in red)

Trajectories always periodic!

What's going on? [Geometric version]

- Dynamic behavior "orthogonal" to potential games   # convergence to Nash
- Is there a dynamic version of the potential/harmonic decomposition?

What's going on? [Geometric version]

- Dynamic behavior "orthogonal" to potential games — # convergence to Nash
- Is there a dynamic version of the potential/harmonic decomposition?

Geometric decomposition of finite games [Legacci et al., 2024]

1. *Any finite game can be decomposed as*

$$v(x) = F(x) + B(x)$$

   where F *is* **irrotational** *and* B *is* **incompressible** *under the Shahshahani metric on* $\mathcal{X}$

2. *A game is* **incompressible** *if and only if it is* **uniform harmonic**

What's going on? [Geometric version]

- Dynamic behavior "orthogonal" to potential games — # convergence to Nash
- Is there a dynamic version of the potential/harmonic decomposition?

Geometric decomposition of finite games [Legacci et al., 2024]

1. *Any finite game can be decomposed as*

$$v(x) = F(x) + B(x)$$

where F is **irrotational** *and B is* **incompressible** *under the Shahshahani metric on $\mathcal{X}$*

2. *A game is* **incompressible** *if and only if it is* **uniform harmonic**

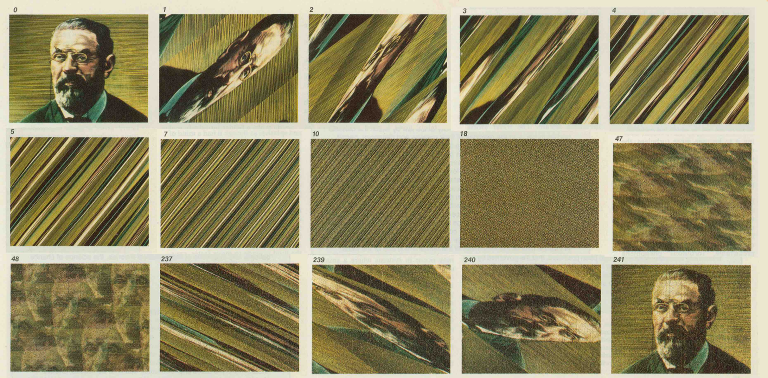
Remarks:

- ***Shahshahani metric*** $\rightsquigarrow$ replicator-compatible geometric structure on $\mathcal{X}$ — # $g_{\alpha\beta}(x) = \delta_{\alpha\beta}/x_\alpha$
- *Why uniform?* — ⚠ highly surprising structural match!

Quasi-periodicity

Poincaré recurrence

A dynamical system is **Poincaré recurrent** if almost all solution trajectories return ***arbitrarily close*** to their starting point ***infinitely many times*** # formal definition of "quasi-periodicity"



Incompressibility, recurrence, and harmonic games

Volume preservation

If Γ is incompressible / harmonic, the replicator dynamics are volume-preserving under the Shahshahani metric:

$$\mathrm{vol}_{\mathrm{Shah}}(A) = \mathrm{vol}_{\mathrm{Shah}}(\mathrm{RD}_t(A)) \quad \text{for every measurable set of initial conditions } A \subseteq \mathcal{X}$$

Incompressibility, recurrence, and harmonic games

Volume preservation

If Γ is incompressible / harmonic, the replicator dynamics are volume-preserving under the Shahshahani metric:

$$\mathrm{vol}_{\mathrm{Shah}}(A) = \mathrm{vol}_{\mathrm{Shah}}(\mathrm{RD}_t(A)) \quad \text{for every measurable set of initial conditions } A \subseteq \mathcal{X}$$

Poincaré recurrence [Legacci et al., 2024]

In any uniform harmonic game, the replicator dynamics are Poincaré recurrent.

Exponential weights redux

Exponential weight dynamics

$$\dot{y}_i = v_i(x) \qquad x_i = \Lambda(y_i) = \frac{\exp(y_i)}{\|\exp(y_i)\|_1} \tag{EWD}$$

Exponential weights redux

Exponential weight dynamics

$$\dot{y}_i = v_i(x) \qquad x_i = \Lambda(y_i) = \frac{\exp(y_i)}{\|\exp(y_i)\|_1} \tag{EWD}$$

Softmax interpretation

$$x = \Lambda(y) \quad \Longleftrightarrow \quad x = \arg\max_{z \in \mathcal{X}} \Big\{ \langle y, z \rangle - \underbrace{\textstyle\sum_{\alpha \in \mathcal{A}} z_\alpha \log z_\alpha}_{\text{entropic penalty}} \Big\}$$

Exponential weights redux

Exponential weight dynamics

$$\dot{y}_i = v_i(x) \qquad x_i = \Lambda(y_i) = \frac{\exp(y_i)}{\|\exp(y_i)\|_1} \tag{EWD}$$

Softmax interpretation

$$x = \Lambda(y) \quad \iff \quad x = \arg\max_{z \in \mathcal{X}} \Big\{ \langle y, z \rangle - \underbrace{\textstyle\sum_{\alpha \in \mathcal{A}} z_\alpha \log z_\alpha}_{\text{entropic penalty}} \Big\}$$

Is there a general principle in play?

Regularized learning

Replace the "**leader**" / hard arg max

$$x_i(t) = \arg\max_{z_i \in \mathcal{X}} \underbrace{\int_0^t u_i(z_i; x_{-i}(\tau))\, d\tau}_{\text{cumulative payoff}}$$

with a "***regularized leader***" / soft arg max:

$$x_i(t) = \arg\max_{z_i \in \mathcal{X}} \Big\{ \underbrace{\int_0^t u_i(z_i; x_{-i}(\tau))\, d\tau}_{\text{cumulative payoff}} - \underbrace{h_i(z_i)}_{\text{penalty}} \Big\}$$

where $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is a strongly convex ***regularizer*** on $\mathcal{X}_i$

Regularized learning

Replace the "**leader**" / hard arg max

$$x_i(t) = \arg\max_{z_i \in \mathcal{X}} \underbrace{\int_0^t u_i(z_i; x_{-i}(\tau))\, d\tau}_{\text{cumulative payoff}}$$

with a "***regularized leader***" / soft arg max:

$$x_i(t) = \arg\max_{z_i \in \mathcal{X}} \Big\{ \underbrace{\int_0^t u_i(z_i; x_{-i}(\tau))\, d\tau}_{\text{cumulative payoff}} - \underbrace{h_i(z_i)}_{\text{penalty}} \Big\}$$

where $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is a strongly convex ***regularizer*** on $\mathcal{X}_i$

Follow-the-regularized-leader (FTRL)

$$\dot{y}_i(t) = v_i(x(t)) \qquad x_i(t) = Q_i(y_i(t)) \tag{FTRL-D}$$

where

$$Q_i(y_i) = \arg\max_{z_i \in \mathcal{X}_i} \{ \langle y_i, z_i \rangle - h_i(z_i) \}$$ # regularized choice / best response map

Example: Projection dynamics

Euclidean / Ridge regularization

Regularizer:

$$h(x) = \frac{1}{2} \sum_\alpha x_\alpha^2$$

Choice map:

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

Example: Projection dynamics

Euclidean / Ridge regularization

Regularizer:

$$h(x) = \frac{1}{2}\sum_{\alpha} x_{\alpha}^{2}$$

Choice map:

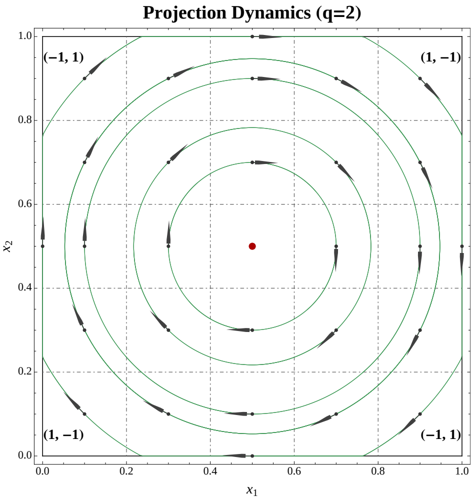
$$\Pi(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - (1/2)\|x\|_2^2\} = \arg\min_{x\in\mathcal{X}}\|y - x\| = \mathrm{proj}_{\mathcal{X}}(y)$$

Projection dynamics [Friedman, 1991; M & Sandholm, 2016]

$$\dot{y}_i(t) = v_i(x(t)) \qquad x_i(t) = \Pi(y_i(t)) \tag{PD}$$

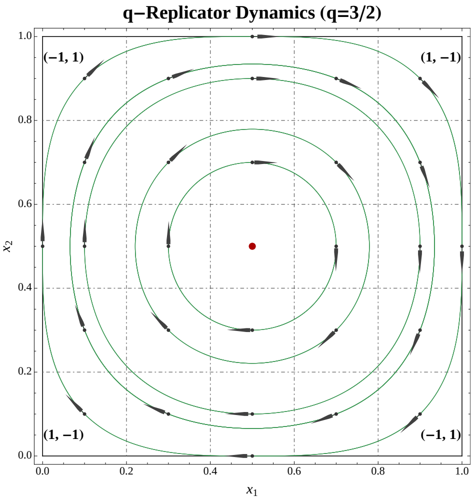
What do the dynamics look like?

The Tsallis kernel: $h(x) = [q(1-q)]^{-1} \sum_\alpha (x_\alpha - x_\alpha^q)$



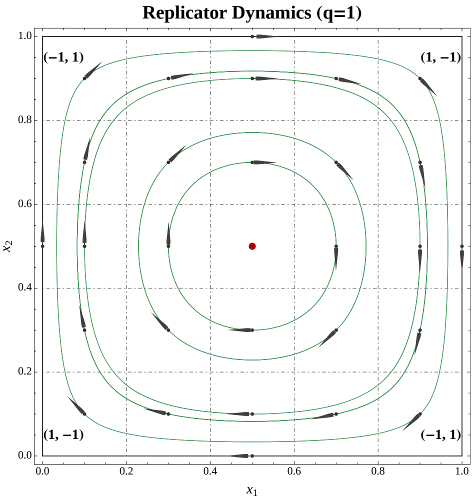
What do the dynamics look like?

The Tsallis kernel: $h(x) = [q(1-q)]^{-1} \sum_{\alpha} (x_{\alpha} - x_{\alpha}^{q})$



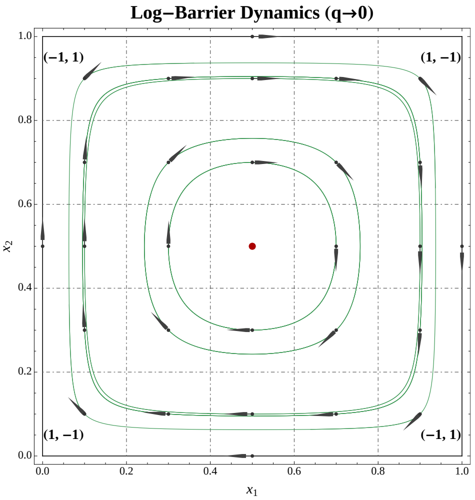
What do the dynamics look like?

The Tsallis kernel: $h(x) = [q(1-q)]^{-1} \sum_\alpha (x_\alpha - x_\alpha^q)$



What do the dynamics look like?

The Tsallis kernel: $h(x) = [q(1-q)]^{-1} \sum_{\alpha} (x_{\alpha} - x_{\alpha}^{q})$



What's going on? [Dual version]

Poincaré recurrence [Legacci et al., 2024]

The dynamics of FTRL are Poincaré recurrent in any harmonic game

What's going on? [Dual version]

Poincaré recurrence [Legacci et al., 2024]

The dynamics of FTRL are Poincaré recurrent in any harmonic game

Remarks:

⚠ **No geometric compatibility** $\implies$ requires completely different proof technique

- Leverage tools from convex analysis $\rightsquigarrow$ constant of motion
- Simultaneously extend to all harmonic measures and all regularizers

Outline

Game-theoretic learning

Sequence of events – discrete time

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

repeat

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

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

Choose ***action*** $\alpha_{i,t} \sim x_{i,t}$ and get ***realized payoff*** $u_i(\alpha_{i,t}; \alpha_{-i,t})$ — # payoff phase

Observe **mixed payoff vector** $v_i(x_t) = (u_i(\alpha_i; x_{-i,t}))_{\alpha_i \in \mathcal{A}_i}$ — # feedback phase

until end

Defining elements

- **Time:** $t = 1, 2, \dots$
- **Players:** many (finite) — # multi-agent learning
- **Actions:** finite
- **Payoffs:** endogenous — # game-theoretic learning
- **Feedback:** **mixed payoff vectors** — # full information, exact

Follow-the-regularized-leader

Follow-the-regularized-leader (FTRL)

$$\begin{aligned} y_{i,t+1} &= y_{i,t} + \gamma_t v_t \\ x_{i,t+1} &= Q_i(y_{i,t+1}) \equiv \arg\max_{x_i \in \mathcal{X}} \{\langle y_{i,t+1}, x_i \rangle - h_i(x_i)\} \end{aligned} \tag{FTRL}$$

- ***Regularized best responses*** instead of logit choice map — M & Sandholm (2016)
- Every player's ***regularizer*** $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is continuous and strongly convex on $\mathcal{X}_i$

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

- **Template includes:** exponential weights, (lazy) projected gradient ascent, Tsallis-based algorithms, ...

Examples

Example (Ridge regularization)

- Regularizer:

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

- Algorithm:

$$y_{t+1} = y_t + \gamma_t v_t \qquad x_{t+1} = \Pi_{\mathcal{X}}(y_{t+1})$$

Example (Entropic regularization)

- Regularizer:

$$h(x) = \sum_{\alpha\in\mathcal{A}} x_\alpha \log x_\alpha$$

- Algorithm:

$$y_{t+1} = y_t + \gamma_t v_t \qquad x_{t+1} = \Lambda(y_{t+1})$$

Non-convergence of FTRL

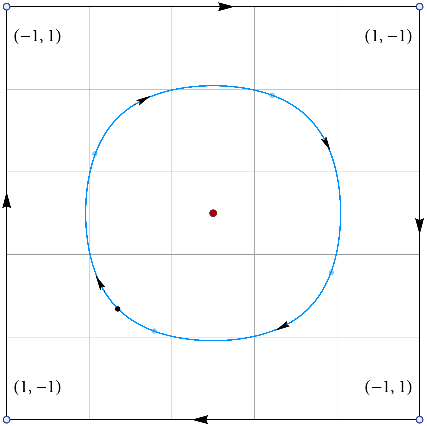


Figure: The replicator dynamics in Matching Pennies

Non-convergence of FTRL

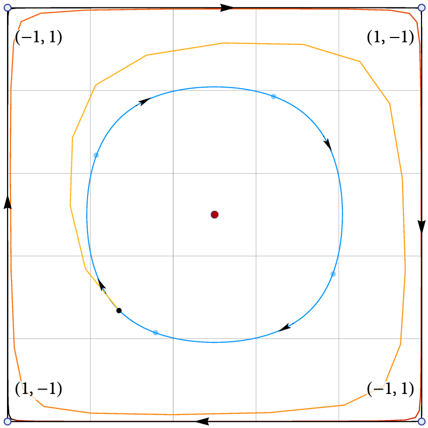


Figure: The FTRL algorithm in Matching Pennies

Non-convergence of FTRL

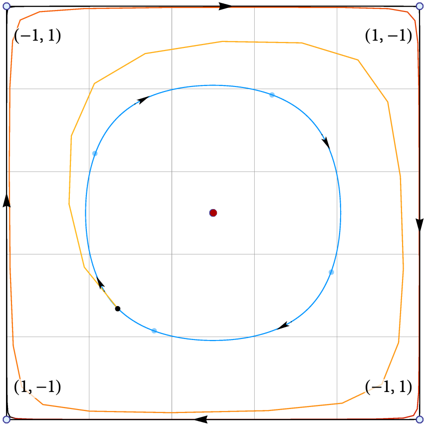


Figure: The FTRL algorithm in Matching Pennies

FTRL does not converge in harmonic games

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

$$\begin{aligned} &a)\ \textit{Extrapolation phase:} && y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} && x_{i,t+1/2} = Q_i(y_{i,t+1/2}) \\ &b)\ \textit{Update phase:} && y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} && x_{i,t} = Q_i(y_{i,t+1}) \end{aligned} \tag{FTRL+}$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

$$\begin{aligned} &a) \text{ Extrapolation phase:} \quad && y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} \quad && x_{i,t+1/2} = Q_i(y_{i,t+1/2}) \\ &b) \text{ Update phase:} \quad && y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} \quad && x_{i,t} = Q_i(y_{i,t+1}) \end{aligned} \tag{FTRL+}$$

Payoff model

$$v_{i,t} =$$

$$v_{i,t+1/2} =$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

a) ***Extrapolation phase:*** $y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} \qquad x_{i,t+1/2} = Q_i(y_{i,t+1/2})$

b) ***Update phase:*** $y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} \qquad x_{i,t} = Q_i(y_{i,t+1})$ (FTRL+)

Payoff model

$$v_{i,t} =$$

$$v_{i,t+1/2} = v_i(x_{t+1/2})$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

$$\begin{aligned}
&a)\ \textit{Extrapolation phase:} \quad && y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} \quad && x_{i,t+1/2} = Q_i(y_{i,t+1/2}) \\
&b)\ \textit{Update phase:} \quad && y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} \quad && x_{i,t} = Q_i(y_{i,t+1})
\end{aligned} \tag{FTRL+}$$

Payoff model

$$\begin{aligned}
v_{i,t} &= 0 \\
v_{i,t+1/2} &= v_i(x_{t+1/2})
\end{aligned}$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

a) ***Extrapolation phase:*** $y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t}$ $\quad x_{i,t+1/2} = Q_i(y_{i,t+1/2})$

b) ***Update phase:*** $y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2}$ $\quad x_{i,t} = Q_i(y_{i,t+1})$ (FTRL+)

Payoff model

$$v_{i,t} = v_i(x_t)$$

$$v_{i,t+1/2} = v_i(x_{t+1/2})$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

$$\begin{aligned} &a) \text{ Extrapolation phase:} && y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} && x_{i,t+1/2} = Q_i(y_{i,t+1/2}) \\ &b) \text{ Update phase:} && y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} && x_{i,t} = Q_i(y_{i,t+1}) \end{aligned} \tag{FTRL+}$$

Payoff model

$$v_{i,t} = v_i(x_{t-1/2})$$
$$v_{i,t+1/2} = v_i(x_{t+1/2})$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

$$\begin{aligned} &a) \textit{ Extrapolation phase:} && y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} && x_{i,t+1/2} = Q_i(y_{i,t+1/2}) \\ &b) \textit{ Update phase:} && y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} && x_{i,t} = Q_i(y_{i,t+1}) \end{aligned} \tag{FTRL+}$$

Payoff model

$$\begin{aligned} v_{i,t} &= \lambda_i v_i(x_t) + (1-\lambda_i) v_i(x_{t-1/2}) \\ v_{i,t+1/2} &= v_i(x_{t+1/2}) \end{aligned}$$

FTRL with an extrapolation step

Extrapolated FTRL (FTRL+)

$$
\begin{aligned}
&a)\ \textit{Extrapolation phase:} \quad && y_{i,t+1/2} = y_{i,t} + \eta_i v_{i,t} \qquad && x_{i,t+1/2} = Q_i(y_{i,t+1/2}) \\
&b)\ \textit{Update phase:} \quad && y_{i,t+1} = y_{i,t} + \eta_i v_{i,t+1/2} \qquad && x_{i,t} = Q_i(y_{i,t+1})
\end{aligned}
\tag{FTRL+}
$$

Payoff model

$$
\begin{aligned}
v_{i,t} &= \lambda_i v_i(x_t) + (1-\lambda_i) v_i(x_{t-1/2}) \\
v_{i,t+1/2} &= v_i(x_{t+1/2})
\end{aligned}
$$

Examples

- $\lambda_i = 1$: Mirror-prox / Extra-gradient — Korpelevich, 1976; Nemirovski, 2004
- $\lambda_i = 0$: Optimistic FTRL / Popov updates — Popov, 1980; ?

Long-run behavior of FTRL+

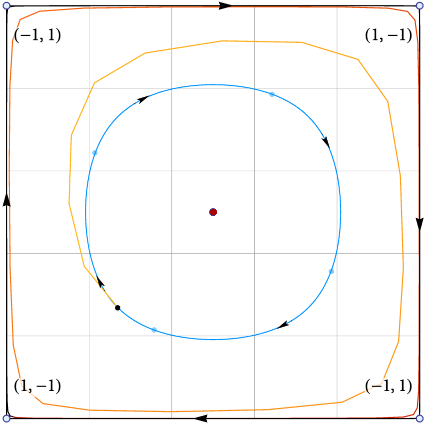


Figure: FTRL in Matching Pennies ✗

Long-run behavior of FTRL+

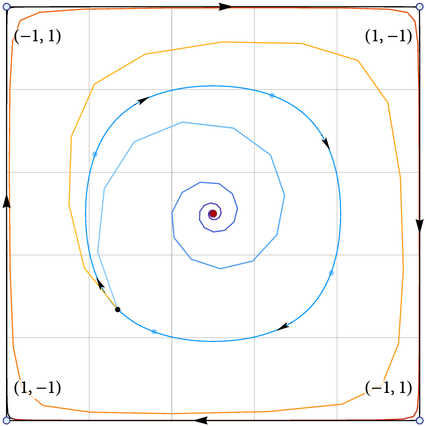


Figure: Mirror-Prox in Matching Pennies ✓

Long-run behavior of FTRL+

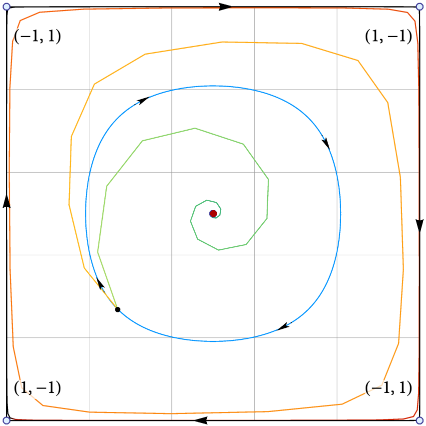


Figure: Optimistic FTRL in Matching Pennies ✓

Long-run behavior of FTRL+

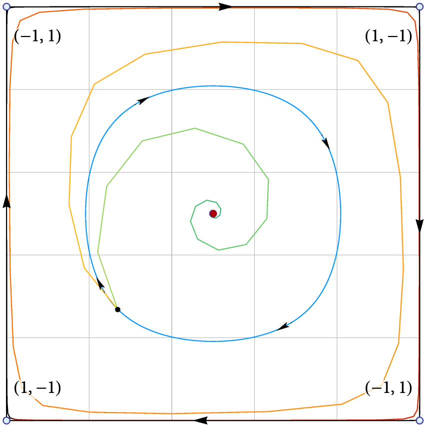


Figure: Optimistic FTRL in Matching Pennies ✓

Does (FTRL+) ***converge in harmonic games?***

Long-run behavior of FTRL+

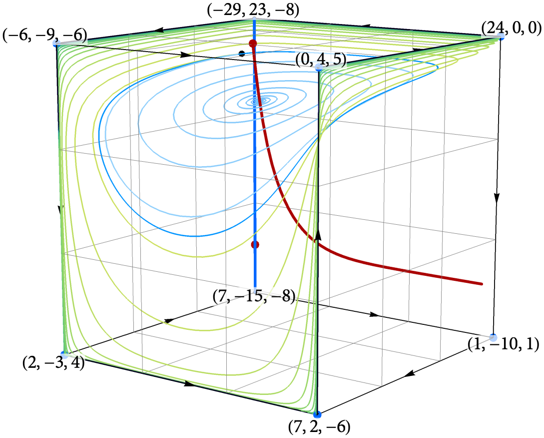


Figure: (FTRL+) in a 2 × 2 × 2 harmonic game ✓

Does (FTRL+) ***converge in harmonic games?***

The long-run behavior of FTRL+

Guarantee 1: Constant regret [Legacci et al., 2024]

Assume:

- Γ is μ-harmonic
- Each player follows (FTRL+) with $\eta_i \leq m_i K_i [2(N+2) \max_j m_j L_j]^{-1}$

Then: (FTRL+) enjoys the bound

$$\operatorname{Reg}_i(T) \leq \frac{H_i}{\eta_i} + \frac{2L_i}{N+2} \sum_{j \in \mathcal{N}} \frac{H_j}{\eta_j L_j} = \mathcal{O}(1)$$

where $H_i = \max h_i - \min h_i$, and L_i is the Lipschitz modulus of v_i

The long-run behavior of FTRL+

Guarantee 1: Constant regret [Legacci et al., 2024]

☞ **Assume:**

- Γ is μ-harmonic
- Each player follows (FTRL+) with $\eta_i \leq m_i K_i [2(N+2) \max_j m_j L_j]^{-1}$

☞ **Then:** (FTRL+) enjoys the bound

$$\mathrm{Reg}_i(T) \leq \frac{H_i}{\eta_i} + \frac{2L_i}{N+2} \sum_{j \in \mathcal{N}} \frac{H_j}{\eta_j L_j} = \mathcal{O}(1)$$

where $H_i = \max h_i - \min h_i$, and L_i is the Lipschitz modulus of v_i

Guarantee 2: Convergence [Legacci et al., 2024]

☞ **Assume:**

- Γ is μ-harmonic
- Each player follows (FTRL+) with $\eta_i \leq m_i K_i [2(N+2) \max_j m_j L_j]^{-1}$

☞ **Then:** the sequence x_t generated by (FTRL+) converges to a Nash equilibrium

Take-aways and conclusions

Main take-aways:

- Harmonic games behave "orthogonally" to potential games in terms of learning
- No-regret learning in continuous time is recurrent
- No-regret learning in discrete time may be divergent...

Take-aways and conclusions

Main take-aways:

- Harmonic games behave "orthogonally" to potential games in terms of learning
- No-regret learning in continuous time is recurrent
- No-regret learning in discrete time may be divergent...
- ...but an extrapolation step recovers convergence to Nash equilibrium

Take-aways and conclusions

Main take-aways:

- Harmonic games behave "orthogonally" to potential games in terms of learning
- No-regret learning in continuous time is recurrent
- No-regret learning in discrete time may be divergent...
- ...but an extrapolation step recovers convergence to Nash equilibrium
- ...and guarantees constant regret

Take-aways and conclusions

Main take-aways:

- Harmonic games behave "orthogonally" to potential games in terms of learning
- No-regret learning in continuous time is recurrent
- No-regret learning in discrete time may be divergent…
- …but an extrapolation step recovers convergence to Nash equilibrium
- …and guarantees constant regret

This is just a first peek:

- Rate of convergence? # difficult, but not hopeless
- Inexact / Payoff-based information # two-step policies?
- Adaptive / Agnostic step-size policies # AdaGrad-like?

References I

[1] Abdou, J., Pnevmatikos, N., Scarsini, M., and Venel, X. Decomposition of games: Some strategic considerations. *Mathematics of Operations Research*, 47(1):176-208, February 2022.

[2] 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.

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

[4] Candogan, O., Menache, I., Ozdaglar, A., and Parrilo, P. A. Flows and decompositions of games: harmonic and potential games. *Mathematics of Operations Research*, 36(3):474-503, 2011.

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

[6] Friedman, D. Evolutionary games in economics. *Econometrica*, 59(3):637-666, 1991.

[7] 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.

[8] 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.

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

[10] Hofbauer, J. and Schlag, K. H. Sophisticated imitation in cyclic games. *Journal of Evolutionary Economics*, 10:523-543, September 2000.

[11] Korpelevich, G. M. The extragradient method for finding saddle points and other problems. *Èkonom. i Mat. Metody*, 12:747-756, 1976.

[12] Lai, T. L. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4-22, 1985.

References II

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

[14] Legacci, D., Mertikopoulos, P., Papadimitriou, C. H., Piliouras, G., and Pradelski, B. S. R. No-regret learning in harmonic games: Extrapolation in the presence of conflicting interests. Preprint, 2024.

[15] Legacci, D., Mertikopoulos, P., and Pradelski, B. S. R. A geometric decomposition of finite games: Convergence vs. recurrence under exponential weights. In *ICML '24: Proceedings of the 41st International Conference on Machine Learning*, 2024.

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

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

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

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

[20] Nemirovski, A. S. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229-251, 2004.

[21] Popov, L. D. A modification of the Arrow-Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR*, 28(5):845-848, 1980.

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

[23] Sorin, S. No-regret algorithms in online learning, games and convex optimization. *Mathematical Programming*, 203:645-686, 2024.

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

[25] Viossat, Y. and Zapechelnyuk, A. No-regret dynamics and fictitious play. *Journal of Economic Theory*, 148(2):825-842, March 2013.