NUS CS6208 - Computational Game Theory

Lecture 8: The Dynamics of Learning in Games

Ryann Sim

PostDoc @ NUS School of Computing

Lesson Plan

- Time-averaged vs day-to-day behavior in no-regret learning (the case for dynamics).
- Divergence of Hedge in zero-sum games.
- Recurrence of replicator dynamics and connecting continuous dynamics with discrete algorithms.
- Optimistic Hedge: Improving regret and ensuring last-iterate convergence in zero-sum games.

Lesson Plan

Recall: No-Regret Learning

At each iteration t, the learner/algorithm selects mixed strategy $x^{(t)} \in \Delta_n$, an adversary selects a loss vector $g^{(t)} \in [0,1]^n$ and the learner obtains reward $\langle x^{(t)}, g^{(t)} \rangle$.

External Regret

$$Reg^T = \max_{x \in \Delta_n} \left\{ \sum_{t=1}^{T} \langle x, g^{(t)} \rangle \right\} - \sum_{t=1}^{T} \langle x^{(t)}, g^{(t)} \rangle$$

Suppose we have a two-player zero-sum game with payoff matrix A. Then, the row player observes utilities $\{-Ay^{(t)}\}_{t=1}^T$ and the column player observes utilities $\{A^\top x^{(t)}\}_{t=1}^T$.

Convergence to NE in Zero-Sum Games

Theorem (Time-Average Convergence)

In two-player zero-sum games with payoff matrix A, if players have regrets Reg_1^T, Reg_2^T, then $(\bar{x}^{(T)}, \bar{y}^{(T)})$ is an ϵ-NE with:

$$\epsilon = \frac{1}{T}(Reg_1^T + Reg_2^T)$$

More broadly, no-regret learning leads to (time-average) convergence to CCE in general-sum normal-form games [Cesa-Bianchi and Lugosi, 2006].

Proof of Time-Average Convergence

Player 1's regret:

$$Reg_1^T = \sum_{t=1}^{T} \langle x^{(t)}, Ay^{(t)} \rangle - T \min_{x'} \langle x', A\bar{y}^{(T)} \rangle \tag{1}$$

Player 2's regret:

$$Reg_2^T = T \max_{y'} \langle y', A^\top \bar{x}^{(T)} \rangle - \sum_{t=1}^{T} \langle x^{(t)}, Ay^{(t)} \rangle \tag{2}$$

Proof of Time-Average Convergence

Player 1's regret:

$$Reg_1^T = \sum_{t=1}^{T} \langle x^{(t)}, Ay^{(t)} \rangle - T \min_{x'} \langle x', A\bar{y}^{(T)} \rangle \tag{1}$$

Player 2's regret:

$$Reg_2^T = T \max_{y'} \langle y', A^\top \bar{x}^{(T)} \rangle - \sum_{t=1}^{T} \langle x^{(t)}, Ay^{(t)} \rangle \tag{2}$$

Summing, we get $Reg_1^T + Reg_2^T = T\left(\max_{y'} \langle y', A^\top \bar{x}^{(T)} \rangle - \min_{x'} \langle x', A\bar{y}^{(T)} \rangle\right)$. The claim follows by summing (1) and (2), and further using the fact that strategy profile (x, y) is an ϵ-NE if $\max_{y'} \langle y', A^\top x \rangle - \min_{x'} \langle x', Ay \rangle \leq \epsilon$.

The Case for Dynamics

What if instead of time-average convergence, we study the day-to-day behavior of learning in games? The core question:

> *In the process of their tâtonnement, will players using a learning rule arrive at an equilibrium state?*

The Case for Dynamics

What if instead of time-average convergence, we study the day-to-day behavior of learning in games? The core question:

> *In the process of their tâtonnement, will players using a learning rule arrive at an equilibrium state?*

Activity 1

	R	P	S
R	$0, 0$	$-1, 1$	$1, -1$
P	$1, -1$	$0, 0$	$-1, 1$
S	$-1, 1$	$1, -1$	$0, 0$

What is the Nash equilibrium of the RPS game above? Suppose two players run the following algorithms: (i) Best-response, (ii) Fictitious play, and (iii) Hedge. Fixing the same (non-NE) initial condition for both players and for each algorithm, compute the first step in each case. What do you observe?

Lesson Plan

1. Algorithmic Performance vs. Dynamical Behavior
2. Divergence of Hedge in Discrete Time
3. On Continuous Dynamics
4. Acceleration and Convergence via Optimism

Hedge in Bimatrix Games

Each player i maintains a weight for each action j: $w_{ij}^{(t)} \in \mathbb{R}$.

1. At each timestep t, each player plays according to the weighted average over all actions.

$$x_{ij}^{(t)} = \frac{w_{ij}^{(t)}}{\sum_j w_{ij}^{(t)}}$$

Hedge in Bimatrix Games

Each player i maintains a weight for each action j: $w_{ij}^{(t)} \in \mathbb{R}$.

1. At each timestep t, each player plays according to the weighted average over all actions.

$$x_{ij}^{(t)} = \frac{w_{ij}^{(t)}}{\sum_j w_{ij}^{(t)}}$$

2. Update weights according to the loss vector and with a stepsize η_i.

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} \exp(-\eta_i \cdot g_{ij}^{(t)})$$

Hedge in Bimatrix Games

Each player i maintains a weight for each action j: $w_{ij}^{(t)} \in \mathbb{R}$.

1. At each timestep t, each player plays according to the weighted average over all actions.

$$x_{ij}^{(t)} = \frac{w_{ij}^{(t)}}{\sum_j w_{ij}^{(t)}}$$

2. Update weights according to the loss vector and with a stepsize η_i.

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} \exp(-\eta_i \cdot g_{ij}^{(t)})$$

We consider a bimatrix game with $G = (A, B)$, player 1 has $u_1(x_1, x_2) = x_1^\top A x_2$ and player 2 has $u_2(x_1, x_2) = x_1^\top B^\top x_2$. Then, $g_{ij}^{(t)}$ is the payoff that would have been obtained by player i if i opts to play pure strategy $j \in \mathcal{A}_i$ when everyone else commits to playing their strategies described by $x := (x_1, x_2)$. We use $v_{ij}(x) := u_i(j; x_{-i})$ to denote this.

Hedge Update Rule

For player i, action j:

$$x_{ij}^{t+1} = \frac{x_{ij}^{t} \exp\left(\eta_i \cdot v_{ij}(x^t)\right)}{\sum_{j' \in \mathcal{A}_i} x_{ij'}^{t} \exp\left(\eta_i \cdot v_{ij'}(x^t)\right)} \tag{Hedge}$$

where $v_{ij}(x^t) = u_i(j; x_{-i}^t)$

Hedge Update Rule

For player i, action j:

$$x_{ij}^{t+1} = \frac{x_{ij}^{t} \exp\left(\eta_i \cdot v_{ij}(x^t)\right)}{\sum_{j' \in \mathcal{A}_i} x_{ij'}^{t} \exp\left(\eta_i \cdot v_{ij'}(x^t)\right)} \tag{Hedge}$$

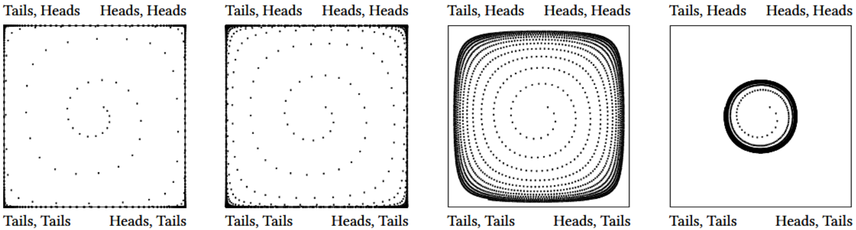
where $v_{ij}(x^t) = u_i(j; x_{-i}^t)$

Example

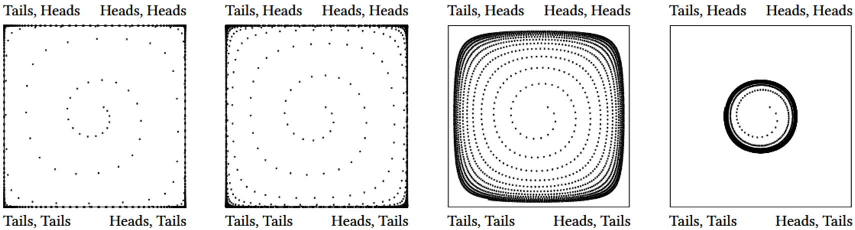
Let's write some simple code to run Hedge on a simple 2 action game, Matching Pennies.

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

Non-Convergence of Hedge



Non-Convergence of Hedge



- Time averages converge to NE
- But actual iterates diverge to boundary!
- Even with decreasing step sizes (from L to R, stepsizes are $0.5, 1/\sqrt[3]{t}, 1/\sqrt{t}, 1/\sqrt[3]{t^2}$).

Non-Convergence of Hedge

KL-Divergence

The KL-divergence between the current iterate x^t and the Nash equilibrium x^* is defined as:

$$D_{KL}(x^* \| x) = \sum_i \sum_j x_{ij}^* (\ln x_{ij}^* - \ln x_{ij})$$

Note that here KL-div is defined as the sum over all players.

Non-Convergence of Hedge

KL-Divergence

The KL-divergence between the current iterate x^t and the Nash equilibrium x^* is defined as:

$$D_{KL}(x^* \| x) = \sum_i \sum_j x^*_{ij}(\ln x^*_{ij} - \ln x_{ij})$$

Note that here KL-div is defined as the sum over all players.

Theorem (Bailey & Piliouras 2018)

In any 2-player zero-sum game with interior NE:

- *KL-divergence is non-decreasing under Hedge*
- *All non-equilibrium initial conditions converge to boundary*
- *Nash equilibria are repelling*

Proof: KL-Divergence is Non-Decreasing I

- Since x_i are distributions on the simplex, we can let $P_i(x) = v_{ij}(x)$ be a random variable which corresponds to the payouts for player i given mixed strategy profile x.
- Only player i introduces randomness to i's reward, and x_{-i} is treated as a deterministic strategy.
- Hence, we can write $u_i(x) = \mathbb{E}[P_i(x)]$.

Proof: KL-Divergence is Non-Decreasing II

Plugging the Hedge update rule into the definition of KL-Div at time $t+1$, we obtain:

$$
\begin{aligned}
D_{KL}(x^*\|x^{t+1}) &= \sum_i \sum_j x^*_{ij}(\ln x^*_{ij} - \ln x^{t+1}_{ij}) && (4)\\
&= \sum_i \sum_j x^*_{ij}\left(\ln x^*_{ij} - \ln x^t_{ij} - \eta_i v_{ij}(x^t) + \ln \mathbb{E}[\exp(\eta_i P_i(x^t))]\right) && (5)\\
&= D_{KL}(x^*\|x^t) + \sum_i \sum_j x^*_{ij}\left(-\eta_i v_{ij}(x^t) + \ln \mathbb{E}[\exp(\eta_i P_i(x^t))]\right) && (6)\\
&= D_{KL}(x^*\|x^t) + \sum_i \left(-\eta_i u_i(x^*_i; x^t_{-i}) + \ln \mathbb{E}[\exp(\eta_i P_i(x^t))]\right) && (7)\\
&= D_{KL}(x^*\|x^t) + \sum_i \left(-\eta_i u_i(x^t) + \ln \mathbb{E}[\exp(\eta_i P_i(x^t))]\right) && (8)
\end{aligned}
$$

Proof: KL-Divergence is Non-Decreasing III

Since the equilibrium x^* is fully mixed, $u_i(x_i^*; x_{-i}^t) = u_i(x^*)$ and moreover since the game is zero-sum, $\sum_i u_i(x_i^*, x_{-i}^t) = \sum_i u_i(x^*) = \sum_i u_i(x^t)$. Thus, from the t to $t+1$-st iteration of Hedge, the KL-divergence changes by:

$$D_{KL}(x^* \| x^{t+1}) - D_{KL}(x^* \| x^t) = \sum_i \left(\ln \mathbb{E}[\exp(\eta_i (P_i(x^t) - u_i(x^t)))] \right) \tag{9}$$

$$\geq \sum_i \eta_i \mathbb{E}[P_i(x^t) - u_i(x^t)] = 0 \tag{10}$$

where the last inequality follows from Jensen's inequality.

Proof Idea: Convergence to Boundary

- KL-divergence is non-decreasing and bounded below.
- Towards a contradiction: Suppose x^{t+1} doesn't converge to boundary $\implies$ $\exists w > 0$ such that $D_{KL}(x^* \| x^{t+1}) \leq w$ for all t.
- Eqn 10 only holds with equality if $x^{t+1} = x^*$ or x^t is on the boundary so, KL-divergence increases by at least $d > 0$ since we initialized away from NE.

$$\lim_{t \to \infty} D_{KL}(x^* \| x^t) = D_{KL}(x^* \| x^0) + \sum_{t=0}^{\infty} \underbrace{(D_{KL}(x^* \| x^{t+1}) - D_{KL}(x^* \| x^t))}_{\geq d} = \infty$$

- This is a contradiction!

Divergence in Games

Extensions

This result is not just for Hedge and not just in two-player zero-sum games! The behavior persists in larger classes of algorithms and games,
see [Bailey and Piliouras, 2018] for more details.

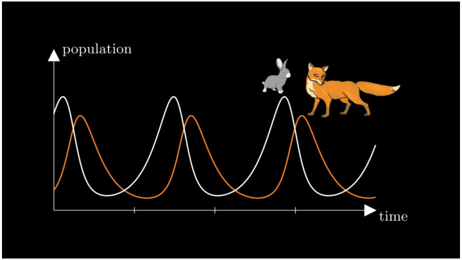
Lesson Plan

On Evolutionary Game Dynamics

- Classical game theory deals with a rational ‘player’, engaged in a ‘game’ with other players and has to choose a ‘strategy’ to maximise a ‘payoff’ which depends on the strategies of the co-players.

On Evolutionary Game Dynamics

- Classical game theory deals with a rational 'player', engaged in a 'game' with other players and has to choose a 'strategy' to maximise a 'payoff' which depends on the strategies of the co-players.
- Evolutionary game theory deals with populations of players, programmed to use some strategy (or type of behaviour).
- Strategies with high payoff will spread within the population (via learning, copying, or even infection).
- Payoffs depend on the actions of the co-players and hence on the frequencies of the strategies within the population.



Replicator Dynamics

Introduced by Taylor & Jonker (1978), replicator dynamics model strategy evolution:

$$\dot{x}_j = x_j(f_j(x) - \phi(x)) \tag{RD}$$

Where:

- x_j: proportion of type j
- $f_j(x)$: fitness of type j
- $\phi(x)$: average population fitness

Replicator Dynamics

Introduced by Taylor & Jonker (1978), replicator dynamics model strategy evolution:

$$\dot{x}_j = x_j(f_j(x) - \phi(x)) \tag{RD}$$

Where:

- x_j: proportion of type j
- $f_j(x)$: fitness of type j
- $\phi(x)$: average population fitness

This model can be translated to bimatrix game (A, B):

$$\dot{x}_{1j} = x_{1j}((Ax_2)_j - x_1^\top Ax_2), \quad j \in \{1, \dots, n\}$$

$$\dot{x}_{2j} = x_{2j}((Bx_1)_j - x_2^\top Bx_1), \quad j \in \{1, \dots, m\}$$

Folk Theorem of Evolutionary Game Theory

Theorem ([Hofbauer and Sigmund, 1998, Cressman and Tao, 2014])

For replicator dynamics, we have the following:

1. *A Nash equilibrium is a rest point*
2. *A stable rest point is a Nash equilibrium*
3. *A convergent interior trajectory converges to a Nash equilibrium*
4. *A strict Nash equilibrium is locally asymptotically stable*

Folk Theorem of Evolutionary Game Theory

Theorem ([Hofbauer and Sigmund, 1998, Cressman and Tao, 2014])

For replicator dynamics, we have the following:

1. *A Nash equilibrium is a rest point*
2. *A stable rest point is a Nash equilibrium*
3. *A convergent interior trajectory converges to a Nash equilibrium*
4. *A strict Nash equilibrium is locally asymptotically stable*

Intuition

- NE: Fixed points ($\dot{x} = 0$)
- Stable rest point: Lyapunov stable + trajectories remain near x^*
- Convergent trajectory: Last-iterate convergence implies NE
- Strict NE: Locally attracting

Folk Theorem of EGT

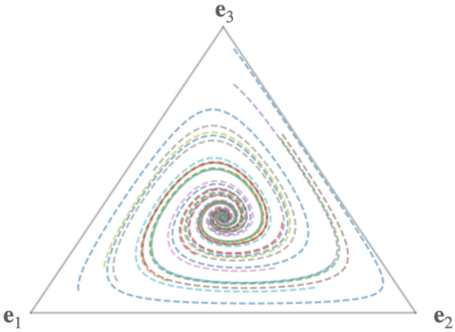


Figure: An example of a monocyclic game, where RD trajectories converge in the interior.

By the folk theorem, we know that the ω-limit point $(1/3, 1/3, 1/3)$ is a Nash equilibrium of the game.

Folk Theorem of EGT

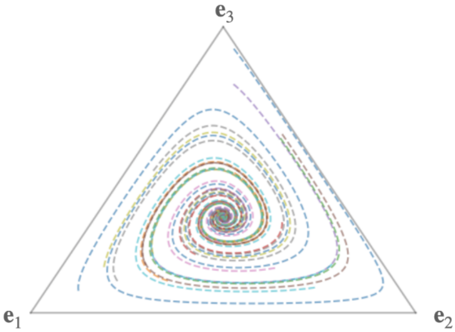


Figure: An example of a monocyclic game, where RD trajectories converge in the interior.

By the folk theorem, we know that the ω-limit point $(1/3, 1/3, 1/3)$ is a Nash equilibrium of the game.

Key Question

Does this behavior always hold if the game has an interior Nash equilibrium?

Recurrence in Zero-Sum Games

Theorem ([Mertikopoulos et al., 2018])

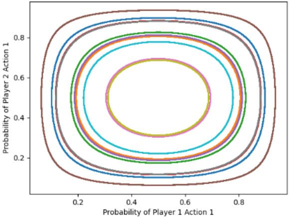
In a 2-player zero-sum game that admits an interior Nash equilibrium, almost every solution trajectory of RD is Poincaré recurrent; specifically, for (Lebesgue) almost every initial condition x^0*, there exists an increasing sequence of times* $t \uparrow \infty$ *such that* $x^t \to x^0$.

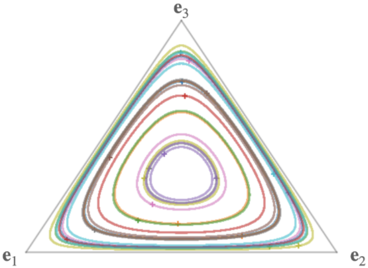
Recurrence in Zero-Sum Games

Theorem ([Mertikopoulos et al., 2018])

In a 2-player zero-sum game that admits an interior Nash equilibrium, almost every solution trajectory of RD is Poincaré recurrent; specifically, for (Lebesgue) almost every initial condition x^0, there exists an increasing sequence of times $t \uparrow \infty$ such that $x^t \to x^0$.

Given any open set of initial conditions, almost all trajectories of RD that start from this set of initial conditions return to it infinitely often. Some games that have this behavior: MP, RPS.





How to show this?

Theorem ([Poincaré, 1890])

If a flow preserves volume and has only bounded orbits then for each open set, almost all orbits intersecting the set intersect it infinitely often.

How to show this?

Theorem ([Poincaré, 1890])

If a flow preserves volume and has only bounded orbits then for each open set, almost all orbits intersecting the set intersect it infinitely often.

1. Show that the volume of the RD system is preserved in zero-sum games with interior NE.
2. Show that the orbits of RD system are bounded in zero-sum games with interior NE.

How to show this?

1. Show that the volume of the RD system (under a smooth, reversible change of variables) is preserved in zero-sum games with interior NE.
2. Show that the orbits of RD system (under a smooth, reversible change of variables) are bounded in zero-sum games with interior NE.

Volume Preservation

Let's illustrate the proof idea with a simple example. Consider the matching pennies game, and focus on the row player, who plays H w.p. x_1 and T w.p. x_2. Replicator dynamics for the first player look like:

$$\frac{\dot{x}_1}{x_1} = u_1 - \sum_{k \in A} x_k u_k, \quad \frac{\dot{x}_2}{x_2} = u_2 - \sum_{k \in A} x_k u_k$$

Volume Preservation

Let's illustrate the proof idea with a simple example. Consider the matching pennies game, and focus on the row player, who plays H w.p. x_1 and T w.p. x_2. Replicator dynamics for the first player look like:

$$\frac{\dot{x}_1}{x_1} = u_1 - \sum_{k \in A} x_k u_k, \quad \frac{\dot{x}_2}{x_2} = u_2 - \sum_{k \in A} x_k u_k$$

Subtracting these, we get:

$$\begin{aligned} \frac{\dot{x}_1}{x_1} - \frac{\dot{x}_2}{x_2} &= u_1 - u_2 \\ \frac{d}{dt} \ln(x_1) - \frac{d}{dt} \ln(x_2) &= u_1 - u_2 \\ \left(\ln\left(\frac{x_1}{x_2}\right)\right)' &= u_1 - u_2 \end{aligned}$$

Volume Preservation

Let's illustrate the proof idea with a simple example. Consider the matching pennies game, and focus on the row player, who plays H w.p. x_1 and T w.p. x_2. Replicator dynamics for the first player look like:

$$\frac{\dot{x}_1}{x_1} = u_1 - \sum_{k \in A} x_k u_k, \quad \frac{\dot{x}_2}{x_2} = u_2 - \sum_{k \in A} x_k u_k$$

Subtracting these, we get:

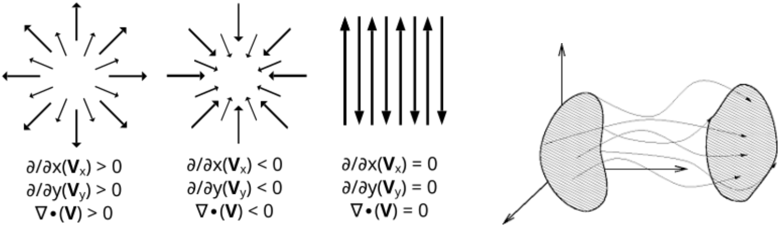
$$\begin{aligned} \frac{\dot{x}_1}{x_1} - \frac{\dot{x}_2}{x_2} &= u_1 - u_2 \\ \frac{d}{dt}\ln(x_1) - \frac{d}{dt}\ln(x_2) &= u_1 - u_2 \\ \left(\ln\left(\frac{x_1}{x_2}\right)\right)' &= u_1 - u_2 \end{aligned}$$

Define new variable $z_1 := \ln\left(\frac{x_1}{x_2}\right) = \ln\left(\frac{x_1}{1-x_1}\right)$. Since $\dot{z}_1 = u_1 - u_2$, we have:

$$z_1 = \int (u_1 - u_2)dt$$

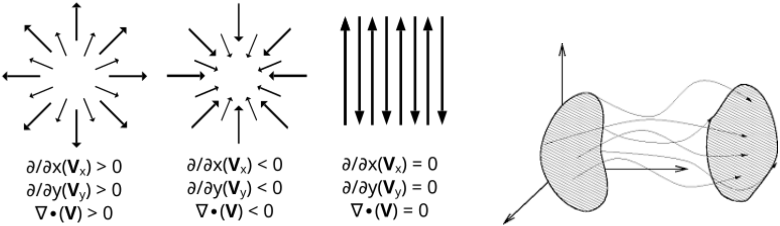
Volume Preservation

Liouville's theorem: If the divergence of a system is zero, then volume is preserved.



Volume Preservation

Liouville's theorem: If the divergence of a system is zero, then volume is preserved.



In our game, and defining z_2 similarly to z_1 we have:

$$\mathrm{Div}(z_1, z_2) = \frac{\partial \dot{z}_1}{\partial z_1} + \frac{\partial \dot{z}_2}{\partial z_2}$$

Why is the divergence zero? Intuitively: $\dot{z}_1$ (instantaneous amount of payoff differences) does NOT depend on what Row player does, but their opponent!

Bounded Orbits

To prove bounded orbits, we derive a *constant of motion*, which in the case of RD is equivalent to the *sum of players' KL-divergences from the NE*!

$$D_{KL}(x^*\|x^0) = D_{KL}(x^*\|x^t) = c, \forall t$$

Bounded Orbits

To prove bounded orbits, we derive a *constant of motion*, which in the case of RD is equivalent to the *sum of players' KL-divergences from the NE*!

$$D_{KL}(x^* \| x^0) = D_{KL}(x^* \| x^t) = c, \forall t$$

- This implies that the dynamics of the cumulative payoff differences do not approach the boundary. Applying Poincaré's theorem completes the proof.
- That is not the only way to show bounded orbits! In fact there is a connection once again to *regret*...

Bounded Orbits

To prove bounded orbits, we derive a *constant of motion*, which in the case of RD is equivalent to the *sum of players' KL-divergences from the NE*!

$$D_{KL}(x^*\|x^0) = D_{KL}(x^*\|x^t) = c, \forall t$$

- This implies that the dynamics of the cumulative payoff differences do not approach the boundary. Applying Poincaré's theorem completes the proof.
- That is not the only way to show bounded orbits! In fact there is a connection once again to *regret*...

Take-Home Activity

Show that replicator dynamics in zero-sum games with interior NE have bounded regret, no matter what other players do. In particular, show that

$$Reg_i^T := \max_{p\in\Delta} \int_0^T u_i(p; x_{-i}(t)) - u_i(x(t))dt \leq \frac{\Omega}{T}$$

where Ω is an absolute constant.

From Replicator to Hedge

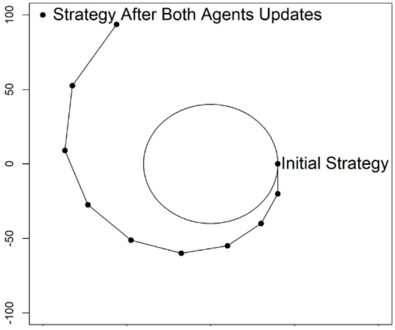
In zero-sum games with interior equilibria, we have some very interesting analogies between the continuous and discrete-time setting for Replicator and Hedge:

- Bounded regret in RD $\rightarrow$ regret growth in Hedge
- Recurrence in RD $\rightarrow$ convergence to boundary in Hedge
- KL-div conserved in RD $\rightarrow$ KL-div increases in Hedge

This is not a coincidence!

Connecting RD and Hedge

The forward Euler discretization of RD is the Hedge update rule.



Different discretization methods of a continuous dynamic leads to different *algorithms* for games!

Beyond Recurrence in Game Dynamics

- Recurrence is **fragile** - breaks with small perturbations.
- Modified RPS (diagonal payoff $\epsilon \neq 0$) exhibits Hamiltonian chaos [Sato et al., 2002].

Beyond Recurrence in Game Dynamics

- Recurrence is **fragile** - breaks with small perturbations.
- Modified RPS (diagonal payoff $\epsilon \neq 0$) exhibits Hamiltonian chaos [Sato et al., 2002].

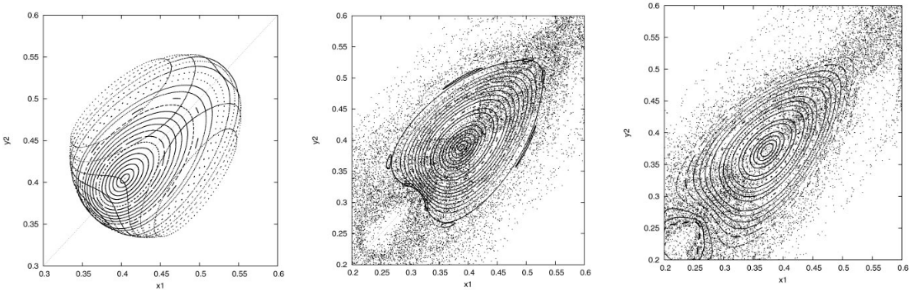
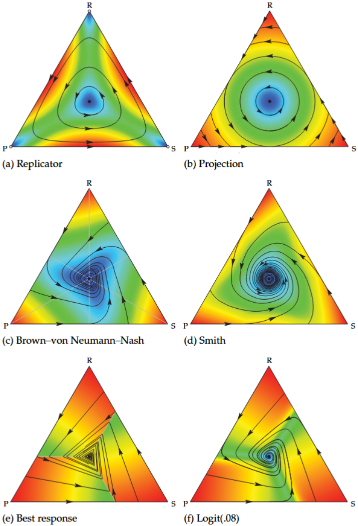


Figure: From L to R, Poincaré section of RD when applied to the RPS game with $\epsilon = 0$, $\epsilon = 0.25$ and $\epsilon = 0.5$.

Different dynamics lead to totally different behaviors! For more info see [Sandholm, 2010].



Impossibility of Universal Convergence

Theorem ([Milionis et al., 2023])

There exist games for which all game dynamics fail to converge to (even approximate) Nash equilibria from all starting points.

Impossibility of Universal Convergence

Theorem ([Milionis et al., 2023])

There exist games for which all game dynamics fail to converge to (even approximate) Nash equilibria from all starting points.

Philosophical Implications

- Is Nash equilibrium the right solution concept for a fundamentally dynamic system?
- Perhaps dynamical behavior itself constitutes the 'solution'.

Lesson Plan

Optimism and Predictivity

Key Insight

Standard Hedge is pessimistic - hedges against worst-case adversaries
But we're playing a static game! Past utilities are predictive.

Optimism and Predictivity

Key Insight

Standard Hedge is pessimistic - hedges against worst-case adversaries
But we're playing a static game! Past utilities are predictive.

- Optimistic algorithms were pioneered in [Rakhlin and Sridharan, 2013] and applied to games later on [Syrgkanis et al., 2015].
- Core idea: At each t, maintain a prediction vector m^t that is updated dynamically, and matches the next loss vector g^t as close as possible.
- Some examples of prediction vectors: average over utils in a fixed window or geometrically discounted past utils.

Optimism and Predictivity

Key Insight

Standard Hedge is pessimistic - hedges against worst-case adversaries
But we're playing a static game! Past utilities are predictive.

- Optimistic algorithms were pioneered in [Rakhlin and Sridharan, 2013] and applied to games later on [Syrgkanis et al., 2015].
- Core idea: At each t, maintain a prediction vector m^t that is updated dynamically, and matches the next loss vector g^t as close as possible.
- Some examples of prediction vectors: average over utils in a fixed window or geometrically discounted past utils.
- We focus on simplest example: $m^t := g^{t-1}$ (one-step recency bias).

Optimistic Hedge

Optimistic Hedge

$$x_{ij}^{t+1} = \frac{x_{ij}^{t} \exp\left(\eta_i \cdot (2v_{ij}(x^t) - v_{ij}(x^{t-1}))\right)}{\sum_{j' \in \mathcal{A}_i} x_{ij'}^{t} \exp\left(\eta_i \cdot (2v_{ij'}(x^t) - v_{ij'}(x^{t-1}))\right)} \qquad \text{(Opt-Hedge)}$$

Optimistic Hedge

Optimistic Hedge

$$x_{ij}^{t+1} = \frac{x_{ij}^{t} \exp\left(\eta_i \cdot \left(2v_{ij}(x^t) - v_{ij}(x^{t-1})\right)\right)}{\sum_{j' \in \mathcal{A}_i} x_{ij'}^{t} \exp\left(\eta_i \cdot \left(2v_{ij'}(x^t) - v_{ij'}(x^{t-1})\right)\right)} \qquad \text{(Opt-Hedge)}$$

- At each iteration, Optimistic Hedge replaces the payoff vector $v_{ij}(x^t)$ with a predictor of the following iteration's payoff vector:

$$v_{ij}(x^t) + (v_{ij}(x^t) - v_{ij}(x^{t-1}))$$

- Two main questions: what is the regret bound and does Opt-Hedge converge in last-iterate to NE?

RVU Bound for Optimistic Hedge

Theorem ([Rakhlin and Sridharan, 2013])

Optimistic Hedge satisfies:

$$Reg^T \leq \alpha + \beta \sum_{t=1}^{T} \|g^t - m^t\|_*^2 - \gamma \sum_{t=1}^{T} \|x^t - x^{t-1}\|^2$$

where $\alpha \propto \frac{1}{\eta}$, $\beta \propto \eta$, $\gamma \propto \frac{1}{\eta}$ *and* $\|\cdot\|_*$ *is the dual norm of* $\|\cdot\|$.

RVU Bound for Optimistic Hedge

Theorem ([Rakhlin and Sridharan, 2013])

Optimistic Hedge satisfies:

$$Reg^T \leq \alpha + \beta \sum_{t=1}^{T} \|g^t - m^t\|_*^2 - \gamma \sum_{t=1}^{T} \|x^t - x^{t-1}\|^2$$

where $\alpha \propto \frac{1}{\eta}$, $\beta \propto \eta$, $\gamma \propto \frac{1}{\eta}$ *and* $\|\cdot\|_*$ *is the dual norm of* $\|\cdot\|$.

- This is known as the Regret Bounded by Variation in Utilities (RVU bound).
- Compare with regret bound of Standard Hedge: $Reg^T \leq \frac{\ln n}{\eta} + \frac{\eta T}{2}$.
- $\sum_{t=1}^{T} \|g^t - m^t\|_*^2$ is the *misprediction error*. In games this is small!
- $-\sum_{t=1}^{T} \|x^t - x^{t-1}\|^2$ implies that large variation in player's strategies $\Longrightarrow$ smaller regret.

Bounded Regret in Zero-Sum Games

Theorem ([Syrgkanis et al., 2015])

With Optimistic Hedge and constant step size η (which is chosen $\leq$ a game dependent constant):

$$Reg_1^T + Reg_2^T \leq \frac{\Omega_1 + \Omega_2}{\eta}$$

This implies $O(1/T)$ convergence to Nash equilibrium.

Bounded Regret in Zero-Sum Games

Theorem ([Syrgkanis et al., 2015])

With Optimistic Hedge and constant step size η (which is chosen $\leq$ a game dependent constant):

$$Reg_1^T + Reg_2^T \leq \frac{\Omega_1 + \Omega_2}{\eta}$$

This implies $O(1/T)$ convergence to Nash equilibrium.

In a zero-sum game, the predictive error term $\sum_{t=1}^{T} \|g^t - m^t\|_*^2$ has a nice representation: For player 1, since $g^t = Ax_2^t$ and $m^t = Ax_2^{t-1}$, we get $\|g^t - m^t\|_*^2 = \|A(x_2^t - x_2^{t-1})\|_*^2$, and similarly for player 2.
Plugging into the RVU bound (Player 1):

$$Reg_1^T \leq \frac{\Omega_1}{\eta} + \eta \sum_{t=1}^{T} \|A(x_2^t - x_2^{t-1})\|_*^2 - \frac{c_1}{\eta} \sum_{t=1}^{T} \|x_1^t - x_1^{t-1}\|^2$$

Proof Sketch

The statement follows from summing up the RVU bounds, and observing that with appropriate choice of η, the middle (predictive error) terms cancel out with the rightmost (variation in utility) terms.

Proof Sketch

The statement follows from summing up the RVU bounds, and observing that with appropriate choice of η, the middle (predictive error) terms cancel out with the rightmost (variation in utility) terms. **Player 1's RVU bound:**

$$Reg_1^T \leq \frac{\Omega_1}{\eta} + \eta \|A\|_{op} \sum_{t=1}^{T} \|x_2^t - x_2^{t-1}\|_*^2 - \frac{c_1}{\eta} \sum_{t=1}^{T} \|x_1^t - x_1^{t-1}\|^2$$

Player 2's RVU bound:

$$Reg_2^T \leq \frac{\Omega_2}{\eta} + \eta \|A\|_{op} \sum_{t=1}^{T} \|x_1^t - x_1^{t-1}\|_*^2 - \frac{c_1}{\eta} \sum_{t=1}^{T} \|x_2^t - x_2^{t-1}\|^2,$$

Selecting the norm appropriately and choosing η carefully, we can get the predictive error and utility variation terms to cancel out!

Bounded Regret in Zero-Sum Games

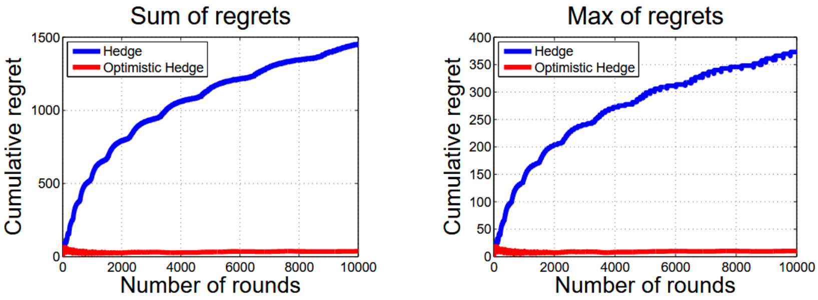


Figure: Comparing regret growth of Hedge and Opt-Hedge in a two-player zero-sum game. Figure taken from [Syrgkanis et al., 2015].

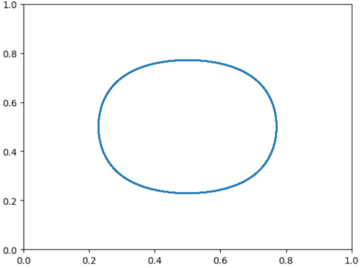
CCE Convergence Rates in General-Sum Games

Opt-Hedge converges faster than Hedge to CCE in general-sum games!

- [Syrgkanis et al., 2015] originally showed a $O(n \log(|A|) T^{-3/4})$ convergence rate to CCE using the RVU bounds.
- [Chen and Peng, 2020] improved this rate to $O(n \log^{5/6}(|A|) T^{-5/6})$ for two-player general-sum normal-form games only.
- [Daskalakis et al., 2021] showed a $O(n \log(|A|) \frac{\log^4 T}{T})$ rate using an extremely complicated technique.
- [Farina et al., 2022] showed a $O(n \log(|A|) \frac{\log T}{T})$ rate using a variant of Opt-Hedge with a different, self-concordant regularizer.

Day-to-Day Behavior of Opt-Hedge

Recall the day-to-day behavior of Hedge vs Replicator:



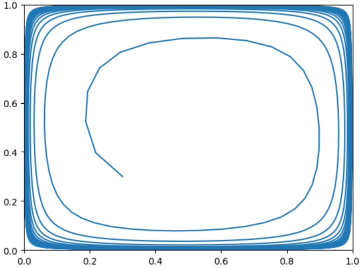


Figure: Replicator vs Hedge: recurrence vs divergence to boundary

We can study Opt-Hedge from this perspective too!

Theorem ([Daskalakis and Panageas, 2019])

In zero-sum games with unique interior Nash:

- *Optimistic Hedge converges in last-iterate*
- *KL-divergence decreases each iteration*
- *Local asymptotic stability around Nash*

Proof Sketch.

1. KL-divergence from t-th iterate to unique NE decreases by $\Omega(\eta^3)$ per iteration, unless the iterate is $O(\eta^{1/3})$-close.
2. If the stepsize η is small enough, then (x^t, y^t) lies in a neighbourhood of (x^*, y^*) that grows smaller as $\eta \to 0$.
3. Once Opt-Hedge reaches a local neighbourhood around (x^*, y^*), the dynamic is locally (asymptotically) stable.

Comparison

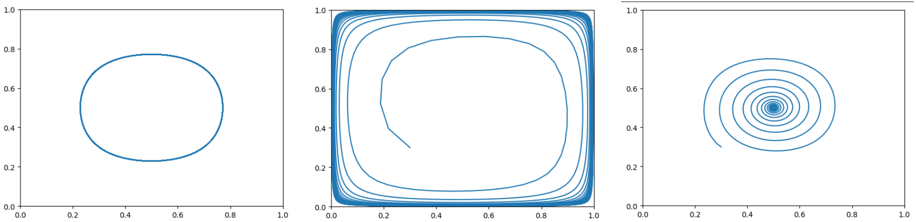


Figure: From L to R, the dynamics of Replicator Dynamics, Hedge and Optimistic Hedge when applied to a Matching Pennies game.

The General Case?

It is natural to ask the following question: *Under what conditions do optimistic no-regret dynamics converge in the last-iterate sense to NE in general games?*

- Opt-Hedge guarantees asymptotic (i.e. in the limit as $t \to \infty$) last-iterate convergence to the set of NE in any two-player zero-sum game [Hsieh et al., 2021].
- Surprisingly, [Cai et al., 2024] showed that even in two-player zero-sum games, the last iterate convergence rate depends on a condition number of the game, and could thus be arbitrarily slow.

Key Takeaways

- Time-average convergence $\neq$ day-to-day behavior.
- Hedge diverges from NE in zero-sum games with interior NE.
- Replicator Dynamics exhibit nicer recurrent behavior in zero-sum games with interior NE, but can be chaotic otherwise.
- Optimistic Hedge enables both better regret bounds and last-iterate convergence in zero-sum games.
- Many exciting directions to understand dynamics in games better + enable faster equilibrium computation!

Questions?

References I

Bailey, J. P. and Piliouras, G. (2018).
Multiplicative weights update in zero-sum games.
In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 321–338.

Cai, Y., Farina, G., Grand-Clément, J., Kroer, C., Lee, C.-W., Luo, H., and Zheng, W. (2024).
Fast last-iterate convergence of learning in games requires forgetful algorithms.
Advances in Neural Information Processing Systems, 37:23406–23434.

Cesa-Bianchi, N. and Lugosi, G. (2006).
Prediction, learning, and games.
Cambridge university press.

Chen, X. and Peng, B. (2020).
Hedging in games: Faster convergence of external and swap regrets.
Advances in Neural Information Processing Systems, 33:18990–18999.

References II

Cressman, R. and Tao, Y. (2014).
The replicator equation and other game dynamics.
Proceedings of the National Academy of Sciences, 111(supplement_3):10810–10817.

Daskalakis, C., Fishelson, M., and Golowich, N. (2021).
Near-optimal no-regret learning in general games.
Advances in Neural Information Processing Systems, 34:27604–27616.

Daskalakis, C. and Panageas, I. (2019).
Last-iterate convergence: Zero-sum games and constrained min-max optimization.
In Blum, A., editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 27:1–27:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

References III

Farina, G., Anagnostides, I., Luo, H., Lee, C.-W., Kroer, C., and Sandholm, T. (2022).
Near-optimal no-regret learning dynamics for general convex games.
Advances in Neural Information Processing Systems, 35:39076–39089.

Hofbauer, J. and Sigmund, K. (1998).
Evolutionary games and population dynamics.
Cambridge university press.

Hsieh, Y.-G., Antonakopoulos, K., and Mertikopoulos, P. (2021).
Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium.
In *Conference on Learning Theory*, pages 2388–2422. PMLR.

Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. (2018).
Cycles in adversarial regularized learning.
In *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*, pages 2703–2717. SIAM.

References IV

Milionis, J., Papadimitriou, C., Piliouras, G., and Spendlove, K. (2023).
An impossibility theorem in game dynamics.
Proceedings of the National Academy of Sciences, 120(41):e2305349120.

Poincaré, H. (1890).
Sur le problème des trois corps et les équations de la dynamique.
Acta mathematica, 13(1):A3–A270.

Rakhlin, A. and Sridharan, K. (2013).
Optimization, learning, and games with predictable sequences.
In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013*, pages 3066–3074.

Sandholm, W. H. (2010).
Population games and evolutionary dynamics.
MIT press.

References V

Sato, Y., Akiyama, E., and Farmer, J. D. (2002).
Chaos in learning a simple two-person game.
Proceedings of the National Academy of Sciences, 99(7):4748–4751.

Syrgkanis, V., Agarwal, A., Luo, H., and Schapire, R. E. (2015).
Fast convergence of regularized learning in games.
Advances in Neural Information Processing Systems, 28.