Machine Learning in Games and Economic Systems

Yun Kuen CHEUNG (Marco)
School of Computing, ANU
21 March, 2024

About Myself

- Theoretical researcher in
  - game theory and market theory
  - algorithm design & analysis
  - combinatorics, graph theory & algorithms

About Myself

- Theoretical researcher in
  - algorithmic game theory and market theory

very inter-disciplinary

dynamic interactions in games & markets	$\Longleftrightarrow$	ML, AI, optimization, & dynamical systems

Learning-in-Games

Learning-in-Games

- Part 1: Motivating Examples
- Part 2: Games & Nash Equilibrium
- Part 3: Online Learning Algorithms
- Part 4: Result Highlights via Dynamical Systems & Optimization

Part 1

Motivating Examples of Learning-in-Games

Machine Learning in Games

What is ML?

insight

data

$$\{\mathbf{x}_i\}_{i=1}^N \sim \mathcal{D}$$

or

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^N \sim \mathcal{D}$$

training →

regression
classification
density estimation
clustering
principal components

The insight is used to make prediction about the future.

Machine Learning in Games

The insight is used to make prediction about the future, which is used for decision making.

$$\{\mathbf{x}_i\}_{i=1}^N \sim \mathcal{D}$$

or

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^N \sim \mathcal{D}$$

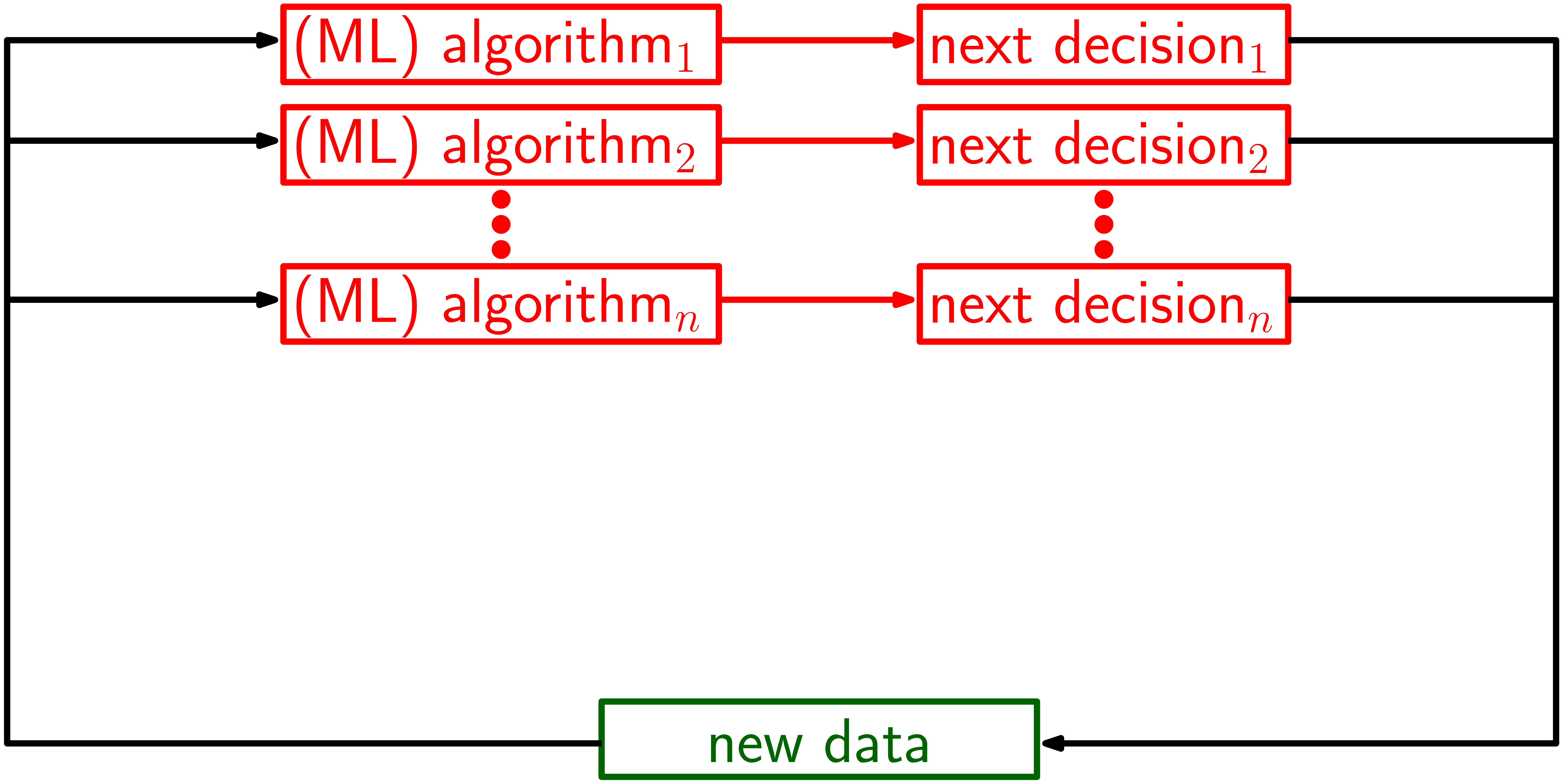
training →

regression
classification
density estimation
clustering
principal components

What if $\mathcal{D}$ is consistently affected by the decisions made using your own and also other (ML) algorithms?

Machine Learning in Games (LiG) is a dynamical system!

Feedback Loop of LiG



Example: Recommender System

- **Platform**:
  - input: users' responses, users' types (e.g., watch history, age, gender)
  - decision: recommendations of videos (e.g., ranking)
  - objective: reach rate / ad revenue / diversity of video types
- **Content creators:**
  - input: users' responses, revenue history
  - decision: types of videos to create
  - objective: revenue, attention, (political) influence

Example: Stock Market

- **Traders**:
  - input: stock price data, news
  - decision: call/put
  - objective: profit
- As more and more big financial companies use algorithmic trading, **regulatory / public / academic** want to know:
  - Can automatic interactions of (ML) algorithms lead to financial instability (even if no breaking news)?
  - Which (ML) algorithm can lead to higher profit and/or less risk?

Example: Adversarial Learning

- **Adversarial Attack** [Szegedy et al.; ICLR 2014]
  - "While DNN's expressiveness is the reason they succeed, it also causes them to learn uninterpretable solutions that have counter-intuitive properties."
  - "DNN learn input-output mappings that are fairly discontinuous... We can cause it to misclassify an image by applying some inperceptible perturbation."
  - Can be viewed as zero-sum game between learner and adversary.

learner adversary

$$\min_{\theta} \max_{\substack{\rho_1,\dots,\rho_N \\ \|\rho_i\| \leq \epsilon}} \mathcal{L}(\theta, \mathbf{X}, \mathbf{y}, \boldsymbol{\rho}) = \sum_{i=1}^{N} \mathsf{KL}(y_i \parallel p_\theta(\mathbf{x}_i + \boldsymbol{\rho}_i))$$

Example: Adversarial Learning

learner adversary

$$\min_{\theta} \max_{\substack{\rho_1,\ldots,\rho_N \\ \|\rho_i\|\leq\epsilon}} \mathcal{L}(\theta, \mathbf{X}, \mathbf{y}, \boldsymbol{\rho}) = \sum_{i=1}^{N} \mathsf{KL}(y_i \parallel p_\theta(\mathbf{x}_i + \boldsymbol{\rho}_i))$$

An effective but less powerful adversary [Goodfellow et al; ICLR 2015]

$$\min_{\theta} \alpha \cdot \mathcal{L}(\theta, \mathbf{X}, \mathbf{y}, \mathbf{0}) + (1 - \alpha) \cdot \mathcal{L}(\theta, \mathbf{X}, \mathbf{y}, \boldsymbol{\rho}) ,$$

where $\rho_i = \epsilon \cdot \mathsf{sign}(\nabla_{\mathbf{x}_i} \mathcal{L}(\theta, \mathbf{X}, \mathbf{y}, \mathbf{0}))$

Example: Generative Adversarial Net

- **Generative Adversarial Network** [Goodfellow et al.; NIPS 2014]
  - "A new framework for estimating generative models via an adversarial process — simultaneously train a generative model G and a discriminative model D."
  - D aims to distinguish training data from generated data.
  - G aims to maximize the chance that D making mistake.
  - "This framework corresponds to a minimax two-player (zero-sum) game."

$$\min_{G} \max_{D} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\text{seed}}}[\log(1 - D(G(\mathbf{z})))]$$

Part 2

Games & Nash Equilibrium

Games and Markets

- Games and markets are two major systems in economics. Both involve self-interest agent behaviors.
- Games concern competition (cf. adversary) and cooperation.
  - action of each player
  - pure/mixed strategy of each player
  - payoffs determined by joint actions of all players
- Markets concern resource allocation.
  - demand and supply, often balanced via prices

Games

		Player 2's action		
		R	P	S
Player 1's action	R	(0,0)	(-1,1)	(1,-1)
	P	(1,-1)	(0,0)	(-1,1)
	S	(-1,1)	(1,-1)	(0,0)

action: choice a player can make

mixed strategy: a probability distribution that a player uses
to pick action randomly

expected payoff: in RPS, total payoff always zero, so zero-sum game

Games

Game model is flexible/arbitrary.

		Player 2's action		
		R	P	S
Player 1's action	R	(0,0)	(-1,1)	(10,-15)
	P	(1,-1)	(0,0)	(-1,1)
	S	(-15,10)	(1,-1)	(0,0)

		Player 2's action		
		R	P	S
Player 1's action	R	(0.5,-0.5)	(-1,1)	(1,-1)
	P	(1,-1)	(0.5,-0.5)	(-1,1)
	S	(-1,1)	(1,-1)	(0.5,-0.5)

Nash Equilibrium

joint mixed strategy profile: each player chooses one mixed strategy

Nash equilibrium: a profile such that no player can unilaterally change strategy to obtain better payoff

A joint mixed strategy profile $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is a Nash equilibrium if for any mixed strategy $\mathbf{x}_i'$ of player i,

$$u_i(\mathbf{x}_i', \mathbf{x}_{-i}) \;\leq\; u_i(\mathbf{x}_i, \mathbf{x}_{-i}) \;;$$

in other words,

$$\mathbf{x}_i \;\in\; \arg\max_{\mathbf{x}_i'} \; u_i(\mathbf{x}_i', \mathbf{x}_{-i}) \;.$$

Nash Equilibrium

Asymmetric Matching Pennies:

	H	T
H	(1,1)	(0,0)
T	(0,0)	(3,3)

Three Nash Equilibria:

- Both players pick "H", each payoff is 1.
- Both players pick "T", each payoff is 3.
- Each player picks "H" with probability $\frac{3}{4}$ and "T" with probability $\frac{1}{4}$, each payoff is $\frac{3}{4}$.

Psychology(?) of Nash Equilibrium

Parity Cooperation Game:

- n isolated players. Each player's action is 0 or 1.
- $s =$ sum of actions of all players
- If s is even, each player gets $\$s$.
- If s is odd, each player gets $\$0$.
- There are at least 2^{n-1} Nash equilibria, with payoffs ranging from 0 to $\sim n$. Which are probable outcomes? We can't really tell.

Minimax Theorem

Theorem [John von Neumann; 1928]

For any two-player zero-sum game,

- Nash equilibrium is *essentially* unique;
- payoff to each player at Nash equilibrium is uniquely determined. $\min_{\mathbf{x}_2} \max_{\mathbf{x}_1} \mathbf{x}_1^\top \mathbf{A} \mathbf{x}_2 = \max_{\mathbf{x}_1} \min_{\mathbf{x}_2} \mathbf{x}_1^\top \mathbf{A} \mathbf{x}_2$

Proof idea. By the strong duality of linear programs.

Yoav Freund and Robert Schapire [GEB 2003] presented a surprisingly elementary learning-in-game proof.

Minimax Theorem

- In general, a game can admit multiple Nash equilibria (cf. local minimas of loss function) with vastly different payoffs.
- Minimax Theorem says two-player zero-sum game is more “predictable”.
- However, learning-in-zero-sum-game is not quite so...

Part 3

Online Learning Algorithms

Online Learning

- Minimax Theorem nails two-player zero-sum games.
- But in reality, games are typically not zero-sum.
- Also, games can be played by players who know neither the full game nor their oppoents' moves.
- We use online/adaptive learning algorithm that runs on the fly relying on partial or local information.

Multiplicative Weights Update

- Suppose we know the payoffs to each action in the history.
- If the past cumulative payoff of an action is much higher than those of other actions, we should choose this action more frequently in the future.

1. Set $x_i^0 = \frac{1}{n}$ for each action $i = 1, 2, \ldots, n$.
2. for $t = 1, 2, \ldots, T$ do:
   - Choose an action j with probability x_j^{t-1}.
   - Observe payoffs p_i^t for each action $i = 1, 2, \ldots, n$.
   - Set $x_i^t \propto x_i^{t-1} \cdot \exp(\epsilon p_i^t)$ for each action $i = 1, 2, \ldots, n$.

Multiplicative Weights Update

MWU enjoys no-regret property for any payoffs.

Theorem. If the payoffs in each round are between 0 and 1, by choosing the step-size $\epsilon = \sqrt{\frac{\log n}{T}}$, we have

$$
\begin{aligned}
& \text{(average payoff received in the } T \text{ rounds)} \\
\geq\ & \text{(average payoff received in the } T \text{ rounds} \\
& \qquad \text{if sticking with the best action)} \ - 2\sqrt{\frac{\log n}{T}}\ .
\end{aligned}
$$

tends to 0 if $T \to \infty$

Multiplicative Weights Update

Corollary. [Freund, Schapire; 2003] Minimax Theorem holds.

Proof idea. In a two-player zero-sum game, consider

- Player 1 uses MWU;
- Player 2 is an almighty adversary — always chooses the worst payoff for Player 1 in each round.

Then by the no-regret theorem and the weak duality (whose proofs are elementary), we can prove the equality

$$\min_{\mathbf{x}_2} \max_{\mathbf{x}_1} \ \mathbf{x}_1^\top \mathbf{A} \mathbf{x}_2 \ = \ \max_{\mathbf{x}_1} \min_{\mathbf{x}_2} \ \mathbf{x}_1^\top \mathbf{A} \mathbf{x}_2$$

in a few lines of calculations.

Optimistic MWU

1. Set $x_i^0 = \frac{1}{n}$ for each action $i = 1, 2, \ldots, n$.
2. for $t = 1, 2, \ldots, T$ do:
   - Choose an action j with probability x_j^{t-1}.
   - Observe payoffs p_i^t for each action $i = 1, 2, \ldots, n$.
   - Set $x_i^t \propto x_i^{t-1} \cdot \exp(\epsilon p_i^t)$ for each action $i = 1, 2, \ldots, n$.

$\epsilon(p_i^t + \epsilon(p_i^t - p_i^{t-1}))$

momentum

Multiplicative Weights Update

Version 1:

1. Set $x_i^0 = \frac{1}{n}$ for each action $i = 1, 2, \ldots, n$.
2. for $t = 1, 2, \ldots, T$ do:
   - Choose an action j with probability x_j^{t-1}.
   - Observe payoffs p_i^t for each action $i = 1, 2, \ldots, n$.
   - Set $x_i^t \propto x_i^{t-1} \cdot \exp(\epsilon p_i^t)$ for each action $i = 1, 2, \ldots, n$.

Version 2, the same algorithm but different implementation:

1. Set $W_i^0 = 0$ for each action $i = 1, 2, \ldots, n$.
2. for $t = 1, 2, \ldots, T$ do:
   - Choose an action j with probability $\propto \exp(\epsilon W_i^{t-1})$.
   - Observe payoffs p_i^t for each action $i = 1, 2, \ldots, n$.
   - Set $W_i^t = W_i^{t-1} + p_i^t$ for each action $i = 1, 2, \ldots, n$.

Follow-the-Regularized-Leader

- W_i^{t-1} = cumulative payoff of action i up to time $t-1$

- FTRL:

exploitation: encourages $\mathbf{x}$ be closer to best action

exploration: (Breg ≈ norm) $\mathbf{u}^n$ is uniform encourages $\mathbf{x}$ be closer to $\mathbf{u}^n$

$$\mathbf{x}^t = \arg\max_{\mathbf{x}\in\Delta^n} \overbrace{\sum_{j=1}^{n} x_j W_j^{t-1}} - \overbrace{\frac{1}{\epsilon}\cdot \text{Breg}(\mathbf{u}^n, \mathbf{x})}$$

- MWU is a special case of FTRL, by setting the Bregman divergence to be the KL divergence.

- FTRL also enjoys no-regret property.

Replicator Dynamics and MWU

In evolutionary game theory, competition between species / animals are often modeled as replicator dynamic below, where x_i is population ratio of species i:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i\Big(f_i(x) - \sum_j x_j f_j(x)\Big),$$

where $f_j(x)$ is the fitness of species j when population composition is x.

Proposition. The above replicator dynamic is equivalent to

$$\frac{\mathrm{d}W_i}{\mathrm{d}t} = f_i(x) \text{ , where } x_j = \frac{\exp(W_j)}{\sum_k \exp(W_k)} \text{ .}$$

Corollary. MWU is the forward Euler discretization of the above dynamical system, by viewing $f_i(x)$ as payoff to action i.

Replicator Dynamics and MWU

Proposition. The replicator dynamic is equivalent to

$$\frac{\mathrm{d}W_i}{\mathrm{d}t} = f_i(x)\ , \ \text{where } x_j = \frac{\exp(W_j)}{\sum_k \exp(W_k)}\ .$$

Proof. Let $S := \sum_k \exp(W_k)$. By the chain rule,

$$\begin{aligned}
\frac{\mathrm{d}x_i}{\mathrm{d}t} &= \frac{\exp(W_i)}{S} \cdot \frac{\mathrm{d}W_i}{\mathrm{d}t} - \frac{\exp(W_i)}{S^2} \cdot \sum_k \exp(W_k) \cdot \frac{\mathrm{d}W_k}{\mathrm{d}t} \\
&= x_i \cdot f_i(x) - x_i \sum_k x_k \cdot f_k(x) \\
&= x_i \left[f_i(x) - \sum_k x_k f_k(x) \right]
\end{aligned}$$

Replicator Dynamics and MWU

Corollary. MWU is the forward Euler discretization of an equivalent version of replicator dynamic (RD).

Due to the above corollary, it seems natural to analyze MWU-learning-in-game systems by analyzing its RD-learning-in-game analogues.

This does provide some insight, but they can have very different qualitative behaviors (almost-recurrence vs. chaos)...

Part 4

Result Highlights via Dynamical Systems & Optimization

Part 4(a)

MWU in Zero-Sum Games

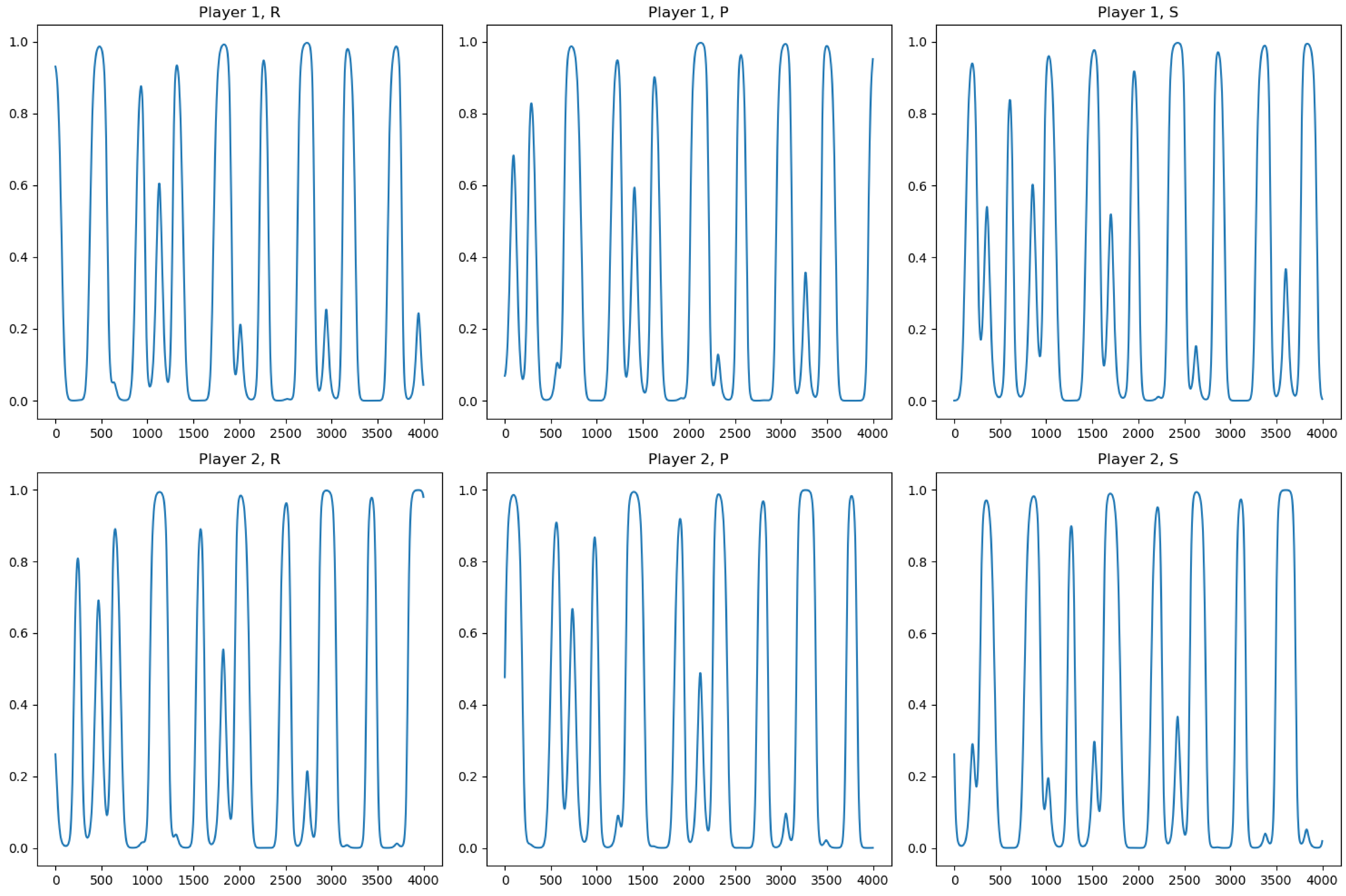
MWU in Zero-Sum Game

- MWU is a natural an popular online learning algorithm.
- Minimax Theorem nails two-person zero-sum game.
- By using the no-regret property of MWU, we have time-average convergence to Nash equilibrium [Freund, Schapire; GEB 2003]:

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_i^t - \mathbf{x}_i^* \right\| \leq \mathcal{O}\left(\frac{1}{\sqrt{T}} \right)$$

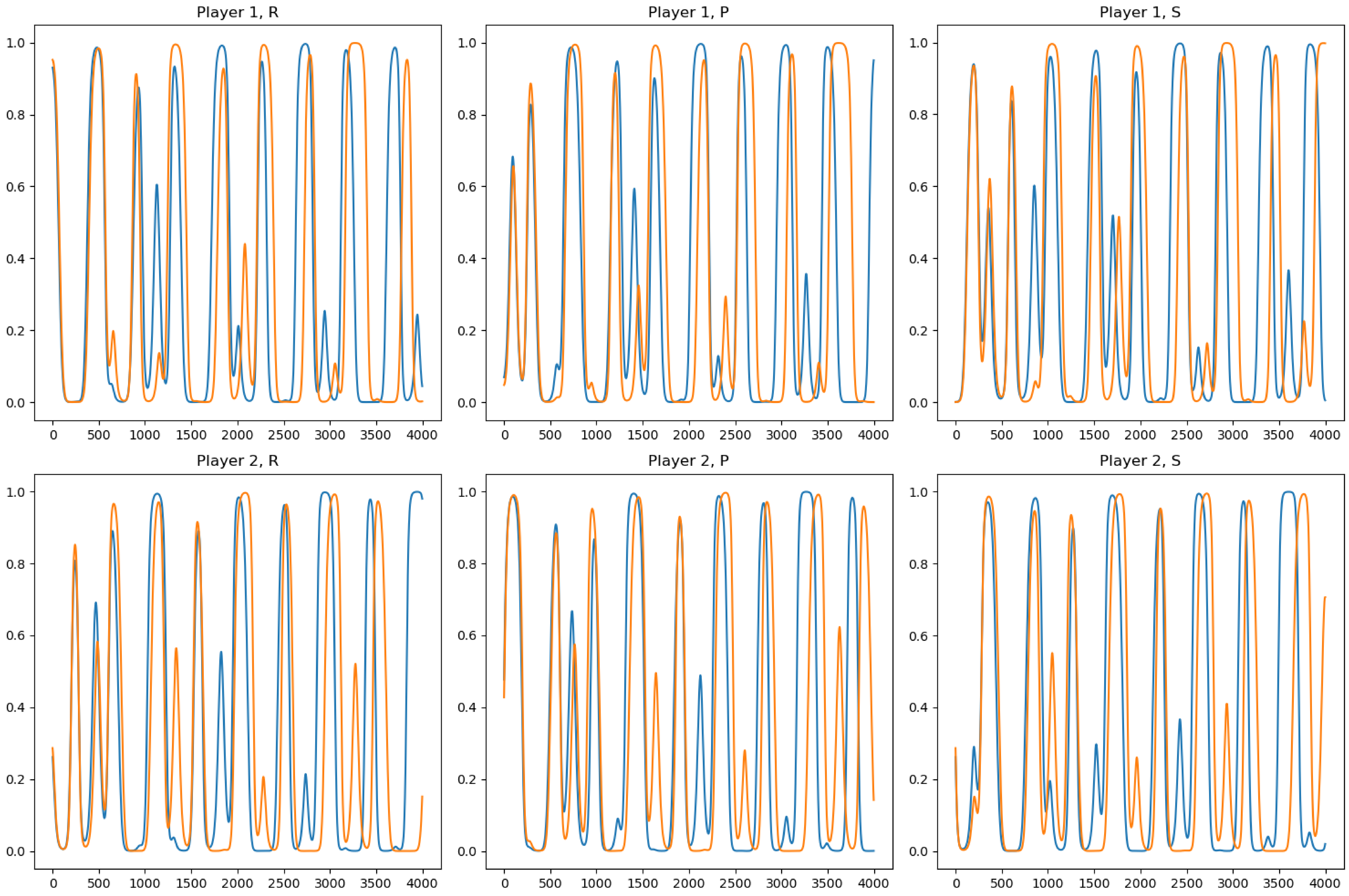
MWU in Zero-Sum Game

But what about the original time series?



MWU in Zero-Sum Game

But what if the starting point perturbs slightly?



MWU in Zero-Sum Game

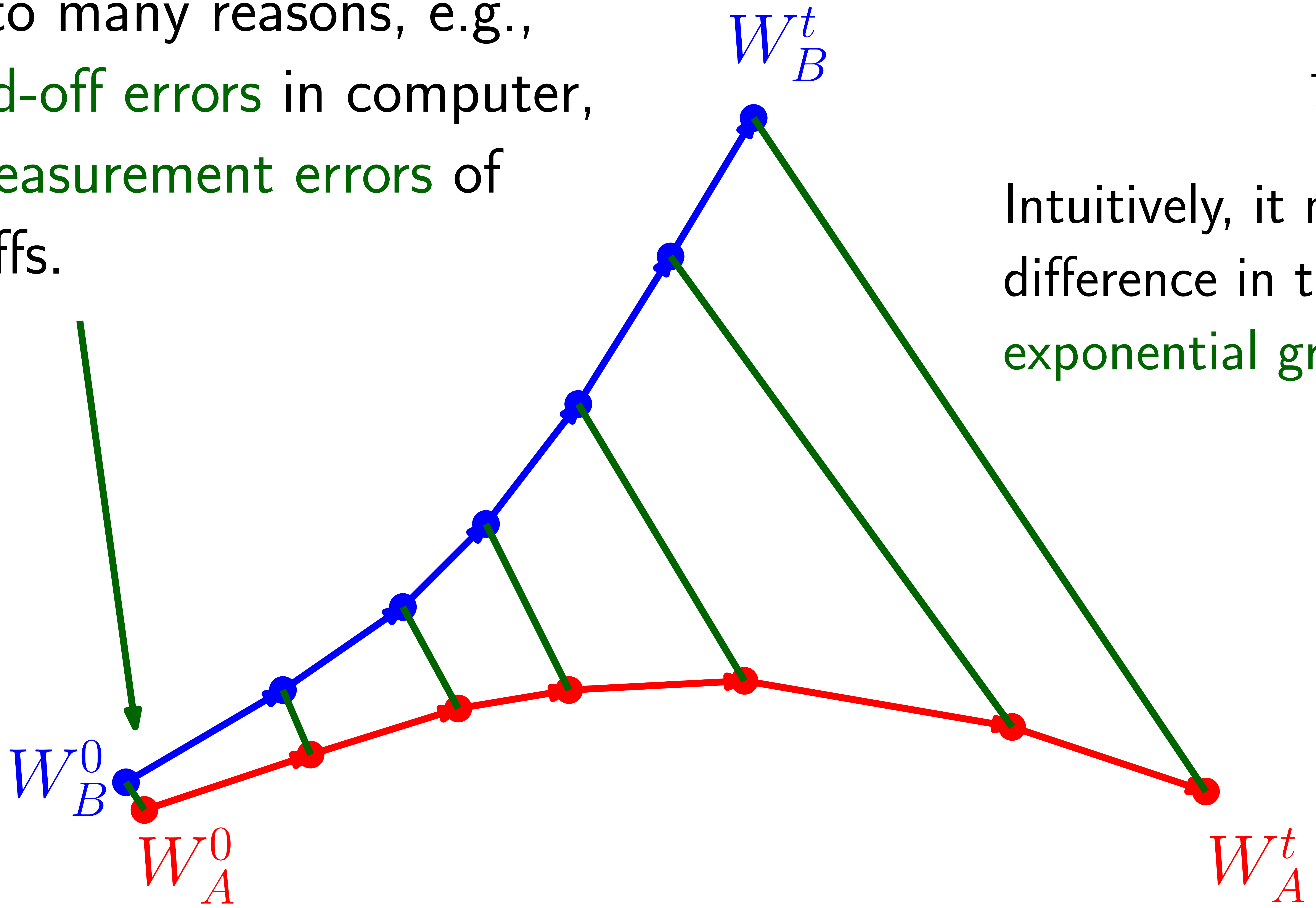
- It seems there is no regular pattern, and even somewhat chaotic.

- Lyapunov chaos (aka butterfly effect) means if a system is initiated with a slightly different state, the trajectories can be very different in the long-term.

  In other words, the system becomes hard to predict in the long run.

Lyapunov Chaos

This perturbation can occur due to many reasons, e.g., round-off errors in computer, or measurement errors of payoffs.



Lyapunov Exponent (informal definition): a lower bound on the maximum possible value of

$$\frac{1}{t} \cdot \log \frac{d(W_A^t, W_B^t)}{d(W_A^0, W_B^0)} .$$

Intuitively, it measures how fast a small difference in the initial condition leads to exponential growth in distance.

Lyapunov Chaos of MWU in Zero-Sum Games

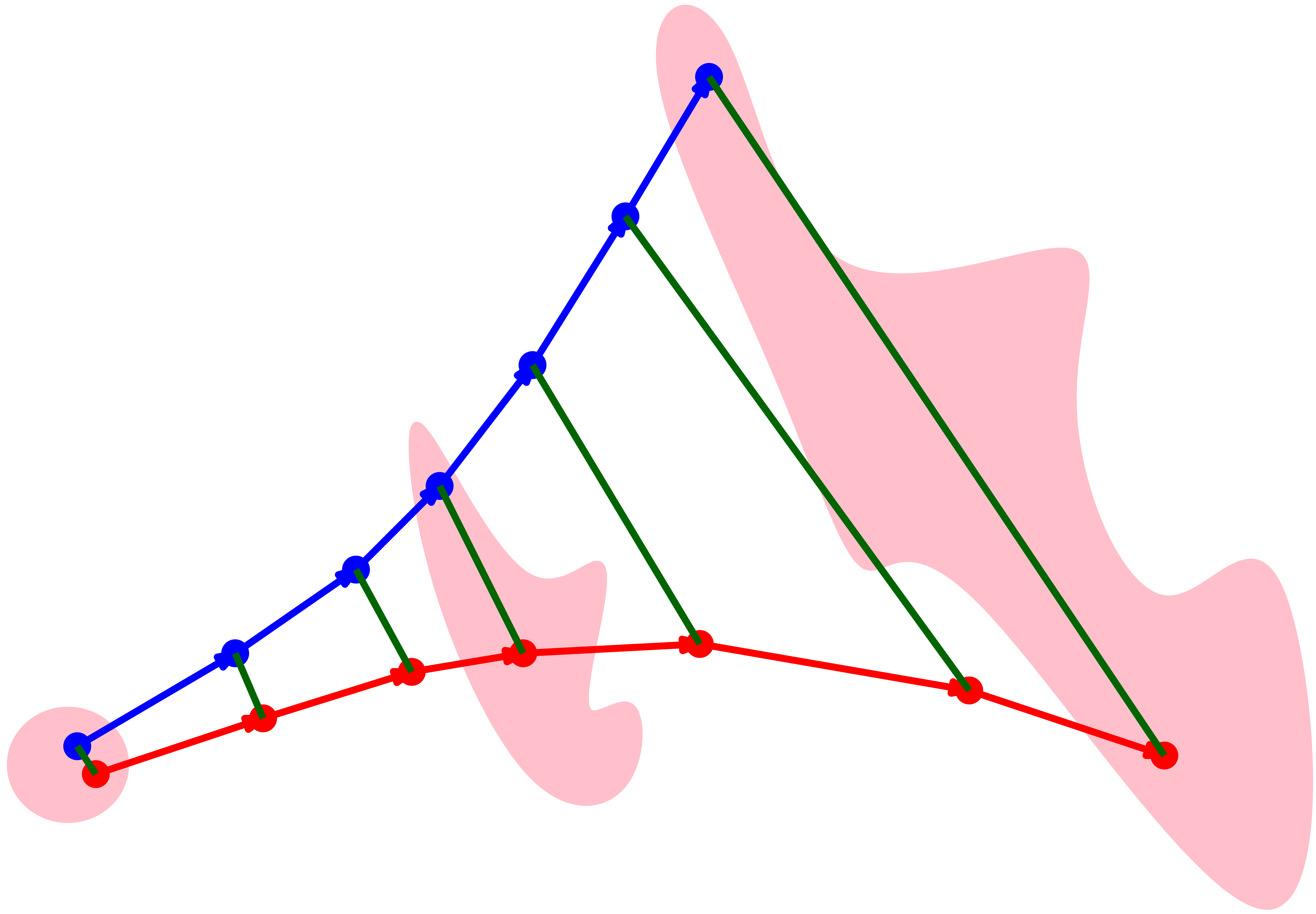
Theorem. [C., Piliouras; COLT 2019]

MWU learning in almost any two-person zero-sum game is globally Lyapunov chaotic in the cumulative payoff space (W-space), with Lyapunov exponent $\Omega(\epsilon^2)$, until the trajectory reaches certain "trivial subspace".

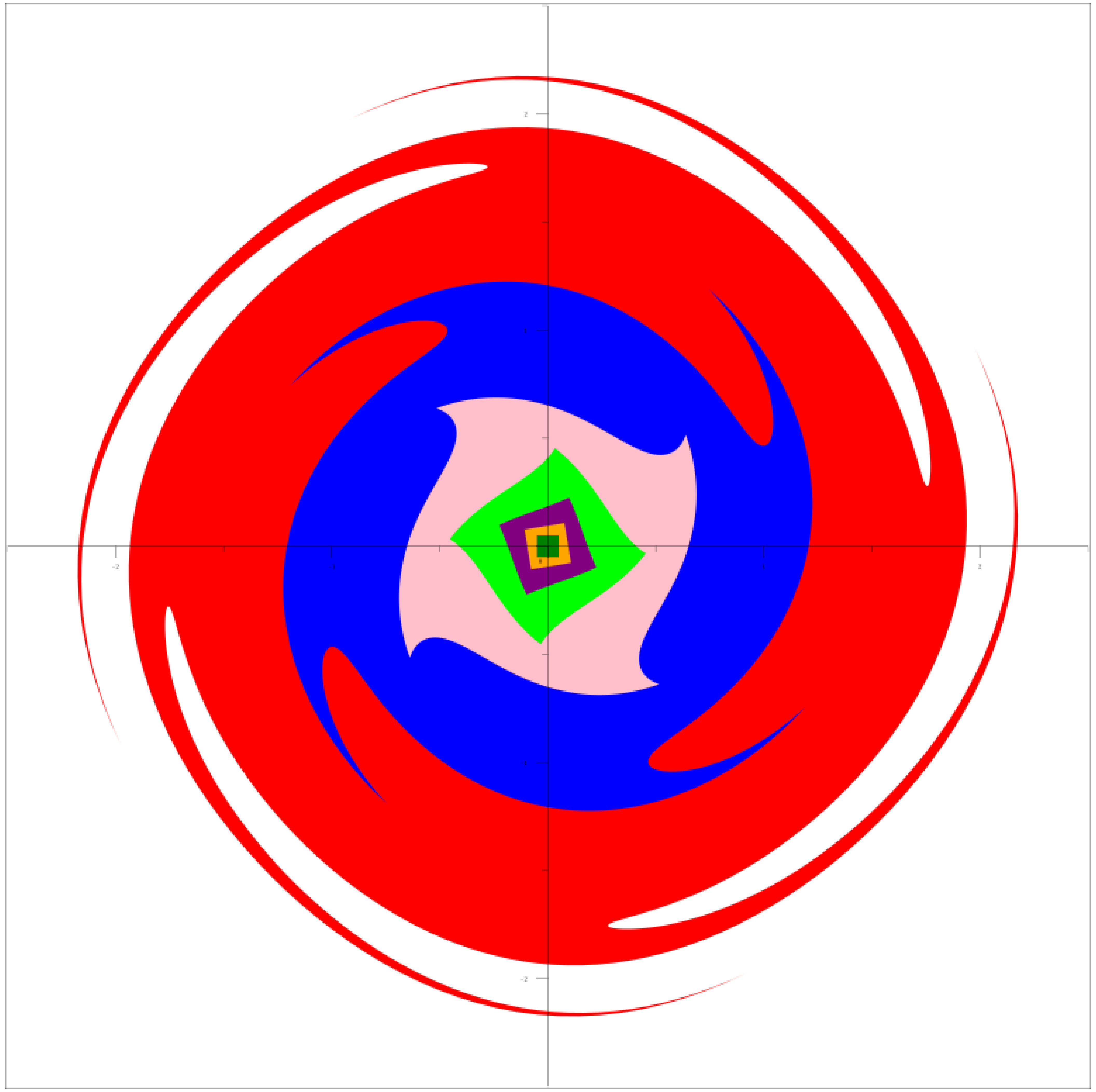
Generalizations:

- FTRL [COLT 2019]
- Graphical constant-sum game [COLT 2019]
- Optimistic MWU in coordination game [C., Piliouras; NeurIPS 2020]
- Direct sum of "strong" zero-sum and "weak" coordination games [C., Tao; ICLR 2021]
- Certain population evolution games [C., Piliouras, Tao; ICLR 2022]

Volume Analysis



Volume Analysis



Volume (in 2D, area) expansion of MWU learning
in Matching Pennies game (which is zero-sum).

Volume Analysis

- Intuitively, when the volume expands exponentially, the diameter also expands at least exponentially.

- For an update rule of the form

$$q_t \leftarrow q_{t-1} + \epsilon \cdot F(q_{t-1}),$$

This integrand decides whether the volume increases or decreases.

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth function. If S is injectively evolved to S' in one time step, then by integration by substitution,

$$\text{volume}(S') = \int_{s \in S} \boxed{\det(\mathbf{I} + \epsilon \cdot \mathbf{J}(s))}\, \mathrm{d}V$$

where $\mathbf{J}(s)$ is the Jacobian matrix of F at s: $J_{ij}(s) = \frac{\partial F_i}{\partial s_j}(s)$.

Part 4(b)

Mirror Descent-Ascent in Convex-Concave Zero-Sum Games

Zero-Sum Game and Minimax

- Two-player zero-sum game can be reformulated as a minimax problem:

$$\min_{\boldsymbol{\theta}_1} \max_{\boldsymbol{\theta}_2} \mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

- This general form includes the adversarial learning framework GAN.
- Wishful thinking: descent on $\boldsymbol{\theta}_1$ and ascent on $\boldsymbol{\theta}_2$, hopefully this process converges to equilibrium.

Zero-Sum Game and Minimax

$$\min_{\boldsymbol{\theta}_1} \max_{\boldsymbol{\theta}_2} \mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

But not that simple in reality...

- Existence of equilibrium?
  - The domains of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ may not be compact.
  - If $\mathcal{L}$ has no global convex-concave structure, convexity of best response correspondence does not hold in general.
  - John Nash's proof on existence of game equilibrium does not work.
- Counter example even on the "simplest" zero-sum games
  - MWU in normal-form zero-sum game $\equiv$ mirror descent-ascent on $\mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_1^\top \mathbf{A} \boldsymbol{\theta}_2$, but we have seen that this is chaotic.

Zero-Sum Game and Minimax

(*Personal opinion*) Not too much hope to establish general and global theoretical results. (Hope someone proves me wrong!)

(Un)fortunately, GAN framework seem working well in practice.

Okay, at least we attempted (unintentionally).

Theorem. [C., Cole, Tao; EC 2018]

If $\mathcal{L}$ is $(\sigma_1, \sigma_2, L_1, L_2)$-strongly Bregman convex-concave, the standard mirror descent-ascent update rule w.r.t. the corresponding Bregman divergence converges to a saddle point (equilibrium), with a linear convergence rate.

Zero-Sum Game and Minimax

Theorem. [Daskalakis, Panageas; ITCS 2019]

When $\mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_1^\top \mathbf{A} \boldsymbol{\theta}_2$, under mild conditions on $\mathbf{A}$, optimistic MWU with a sufficiently small step-size converges to equilibrium.

A non-exhaustive list of interesting relevant work:

- Daskalakis, Panageas, "The Limit Points of (Optimistic) Gradient Descent in Min-Max Optimization", NeurIPS 2018.
- Wang, Zhang, Ba, "On Solving Minimax Optimization Locally", ICLR 2019.
- Daskalakis, Skoulakis, Zampetakis, "The Complexity of Constrained Min-Max Optimization", STOC 2021.
- Daskalakis, Golowich, Skoulakis, Zampetakis, "Guaranteed Convergence to Local Minimax Equilibrium in Nonconvex-Nonconcave Games", COLT 2023.

Summary & Remarks

Summary and Remarks

- Learning-in-Games (LiG) finds many nautural applications in modern (internet) economics.
- The natural association between zero-sum games & minimax optimization is of particular ML (adversary learning) interest.
- Theoretical results are somewhat delicate, depending on the combinations of games and learning algorithms.
- Mathematical insight from both Optimization Theory and Dynamical Systems is necessary for thorough understanding.