

# Learning in Multi-Player Games: Regret, Convergence, and Efficiency

**Ioannis Anagnostides**

Computational Game Solving

(Fall 2023)

# Focus of this lecture

- Closer look at the performance of **no-regret dynamics**
- **Last-iterate convergence**
- **Social welfare** guarantees of no-regret dynamics

# Multi-player games

- Finite number of  $n$  players
- Each player selects a **strategy**  $x_i \in \mathcal{X}_i$
- There is a **utility function**  $u_i : \times_{j=1}^n \mathcal{X}_j \rightarrow \mathbb{R}$
- Once we fix the rest of the players, the **utility function** is linear
- This captures **extensive-form** and **normal-form** games



# The no-regret framework

- A sequence of interactions between a **learner** and the **environment**
- In each round, the learner chooses a **strategy**  $x_i^t$ , and observes a **utility**  $u_i^t$
- Recall the definition of **regret**:

$$\text{Reg}_i^T = \max_{x_i^*} \left\{ \sum_{t=1}^T \langle x_i^*, u_i^t \rangle \right\} - \sum_{t=1}^T \langle x_i^t, u_i^t \rangle$$

# The no-regret framework

- A sequence of interactions between a **learner** and the **environment**
- In each round, the learner chooses a **strategy**  $x_i^t$ , and observes a **utility**  $u_i^t$
- Recall the definition of **regret**:

$$\text{Reg}_i^T = \max_{x_i^*} \left\{ \sum_{t=1}^T \langle x_i^*, u_i^t \rangle \right\} - \sum_{t=1}^T \langle x_i^t, u_i^t \rangle$$

- Regret can be negative!

# The no-regret framework

- A sequence of interactions between a **learner** and the **environment**
- In each round, the learner chooses a **strategy**  $x_i^t$ , and observes a **utility**  $u_i^t$
- Recall the definition of **regret**:

$$\text{Reg}_i^T = \max_{x_i^*} \left\{ \sum_{t=1}^T \langle x_i^*, u_i^t \rangle \right\} - \sum_{t=1}^T \langle x_i^t, u_i^t \rangle$$

- Regret can be negative!
- E.g.,

$$x_i^1 = u_i^1 = (1, 0), x_i^2 = u_i^2 = (0, 1), \dots$$

# No-regret learning in games

- **Each player** updates its strategy via a **no-regret** algorithm
- **Decentralized** and **uncoupled** equilibrium computation
  - Unknown game accessed via **utility queries**
- **Centralized** equilibrium computation
  - State of the art algorithms in theory and in practice

# No-regret learning in games

- Many algorithms (MWU, RM, RM+) guarantee regret at most  $O(\sqrt{T})$
- Convergence to **Nash equilibria** in 2p0s games, and **coarse correlated equilibria** in multi-player general-sum games with a rate of  $T^{-1/2}$
- Are there algorithms that enjoy a **faster rate** of convergence of  $T^{-1}$ ?
- The analysis of  $O(\sqrt{T})$  has been is overly pessimistic
- Here we actually have certain control over the utilities
- Can we improve our analysis? In general, **no!**



# Lower bounds under common regret minimizers

- **Theorem** (Chen-Peng 2020, NeurIPS). *MWU incurs  $\Omega(\sqrt{T})$  regret even in self-play.*
- **Theorem** (Farina-Grand-Clément-Kroer-Lee-Luo 2023, NeurIPS). *RM+ incurs  $\Omega(\sqrt{T})$  regret even in self-play.*

The key technique to obtaining **near-optimal rates** in games revolves around the use of **optimism**.

# Optimistic no-regret learning

- The key idea is to use a **prediction**  $m_i^t$
- Typically set as  $m_i^t = u_i^{t-1}$  (**more sophisticated** predictions?)
- Taking  $m_i^t = 0$  recovers the non-optimistic algorithms
- **Optimistic** FTRL (optimistic MD is defined similarly):

$$x_i^{t+1} = \operatorname{argmax}_{x_i^*} \left\{ \left\langle x_i^*, m_i^{t+1} + \sum_{\tau=1}^t u_i^\tau \right\rangle - \frac{1}{\eta} \mathcal{R}(x_i^*) \right\}.$$

Regularizer  
Learning rate

# Analyzing the regret of optimistic algorithms

**Theorem** (Syrgkanis-Agarwal-Luo-Schapire 2015, NIPS). For any sequence of utilities, the regret of *optimistic FTRL* and *optimistic MD* satisfies

$$\text{Reg}_i^T \leq \frac{\alpha}{\eta} + \beta\eta \sum_{t=1}^T \|u_i^t - m_i^t\|_*^2 - \frac{\gamma}{\eta} \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2. \quad (\text{RVU Bound})$$


The non-optimistic counterparts satisfy 
$$\text{Reg}_i^T \leq \frac{\alpha}{\eta} + \eta \sum_{t=1}^T \|u_i^t\|_*^2.$$

# Analysis of (online) gradient descent

Online gradient descent:  $x_i^{t+1} = \operatorname{argmax}_{x_i^* \in \mathcal{X}_i} \left\{ \langle x_i^*, u_i^t \rangle - \frac{1}{2\eta} \|x_i^* - x_i^t\|_2^2 \right\} \iff x_i^{t+1} = \Pi_{\mathcal{X}_i} (x_i^t + \eta u_i^t).$

Quadratic growth:  $\langle x_i^{t+1}, u_i^t \rangle - \frac{1}{2\eta} \|x_i^{t+1} - x_i^t\|_2^2 - \langle x_i^*, u_i^t \rangle + \frac{1}{2\eta} \|x_i^* - x_i^t\|_2^2 \geq \frac{1}{2\eta} \|x_i^{t+1} - x_i^*\|_2^2.$

$\Phi(x_i^{t+1}) - \Phi(x_i^*) \geq \frac{\mu}{2} \|x_i^{t+1} - x_i^*\|_2^2$



Summing over all time steps,  $\sum_{t=1}^T \langle x_i^* - x_i^t, u_i^t \rangle \leq \frac{1}{2\eta} \|x_i^0 - x_i^*\|_2^2 + \sum_{t=1}^T \langle x_i^{t+1} - x_i^t, u_i^t \rangle.$

# Analyzing the regret of optimistic algorithms

**Lemma.** *If player  $i$  follows (optimistic) MD or FTRL,  $\|x_i^t - x_i^{t-1}\| = O(\eta)$ .*

$\Rightarrow$  If **all** players follow (optimistic) MD or FTRL,  $\|u_i^t - u_i^{t-1}\|_* = O(\eta)$ .

Thus,

$$\text{Reg}_i^T \leq \frac{\alpha}{\eta} + \beta\eta \sum_{t=1}^T \|u_i^t - u_i^{t-1}\|_*^2 \leq \inf_{\eta} \left\{ \frac{\alpha}{\eta} + \beta\eta^3 T \right\} = O(T^{1/4}).$$

# Near-optimal regret in games

- The previous analysis failed to use the last term in the RVU bound

$$\text{Reg}_i^T \leq \frac{\alpha}{\eta} + \beta\eta \sum_{t=1}^T \|u_i^t - m_i^t\|_*^2 - \frac{\gamma}{\eta} \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2.$$

- As a warm-up, we focus on the class of games such that  $\sum_{i=1}^n \text{Reg}_i^T \geq 0$ 
  - Two-player zero-sum games
  - Strategically zero-sum games
  - Polymatrix zero-sum games

$$u_i = \sum_{j \in \mathcal{N}_i} x_i^\top A_{i,j} x_j$$

# Near-optimal regret in games with nonnegative regrets

**Theorem.** *If  $\sum_{i=1}^n \text{Reg}_i^T \geq 0$ , then  $\text{Reg}_i^T = O(1)$ .*

**Proof:** For any **player**  $i$ ,

$$\text{Reg}_i^T \leq \frac{\alpha}{\eta} + \beta' \eta \sum_{j \neq i} \sum_{t=1}^T \|x_j^t - x_j^{t-1}\|^2 - \frac{\gamma}{\eta} \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2.$$

For a sufficiently small **learning rate**,

$$\sum_{i=1}^n \text{Reg}_i^T \leq \frac{\alpha n}{\eta} - \frac{\gamma'}{\eta} \sum_{i=1}^n \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2.$$

Thus,  $\sum_{i=1}^n \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2 = O(1)$ .

# Near-optimal regret in general games

What about general games?

**Theorem** (A-Farina-Luo-Lee-Kroer-Sandholm 2022, NeurIPS). *There exists a no-regret learning algorithm such that for any sequence of utilities,*

$$\max\{\text{Reg}_i^T, 0\} \leq \frac{\alpha \log T}{\eta} + \beta \eta \sum_{t=1}^T \|u_i^t - m_i^t\|_*^2 - \frac{\gamma}{\eta} \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2.$$

- The algorithm is optimistic FTRL with **logarithmic regularization**:  $-\sum_a \log x_i(a)$



# Best of both worlds

- $O(\log T)$  regret is possible when **all** players **follow** the prescribed protocol
- What if some player **deviates**? Can we still secure  $O(\sqrt{T})$  regret?

# Best of both worlds

- $O(\log T)$  regret is possible when **all** players **follow** the prescribed protocol
- What if some player **deviates**? Can we still secure  $O(\sqrt{T})$  regret?

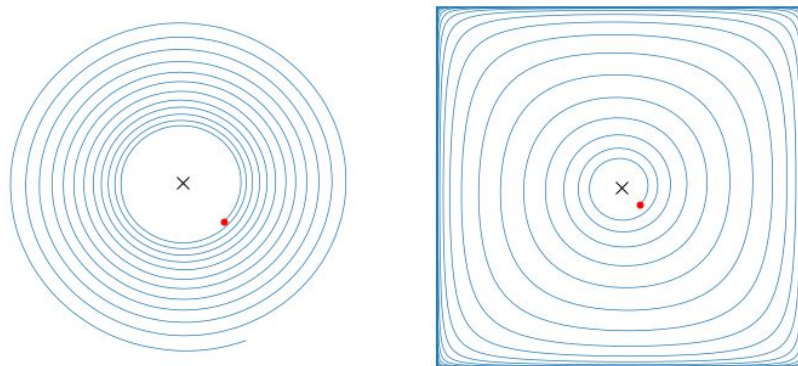
Check whether  $\sum_{\tau=1}^t \|u_i^\tau - u_i^{\tau-1}\|^2 = \Omega(\log t).$

**Yes:** Switch to the **adversarial regime**

**No:** **Follow** the protocol

# Beyond time-average convergence

- The **no-regret** framework implies convergence for the **average strategies**
- What can be said about the **last-iterate** of the dynamics?
- In general, **no-regret** dynamics **cycle** even in 2p0s games



# Importance of last-iterate convergence

- **Algorithmic benefits:** Last-iterate convergence behaves fundamentally different than that of the average iterate
  - Last-iterate can converge at an **exponential rate** (much faster than  $T^{-1}$ ) (Tseng 1995, JCAM; Gilpin-Peña-Sandholm 2012, MathProg; Wei-Lee-Zhang-Luo 2021, ICLR)
  - Only need to store a single strategy (crucial when each strategy is represented with a massive neural network)
- Insights into obtaining **improved regret** guarantees
- A more convincing notion of learning

# Optimism to the rescue

- **Optimistic** learning dynamics have been shown to enjoy **last-iterate convergence** in certain classes of games (e.g. 2p0s)
- **Theorem** ([Wei-Lee-Zhang-Luo 2021, ICLR](#)). *Optimistic gradient descent converges to an  $\epsilon$ -Nash equilibrium in 2p0s games after  $C \log(1/\epsilon)$  iterations.*
  - Main caveat:  $C$  can be arbitrarily large even in  $2 \times 2$  games
  - When is  $C$  small?
  - The limit point is the **projection** of the initialization to the set of NE!
  - Last-iterate is an extreme version of **weighted averages**

# Analyzing last-iterate convergence

We saw earlier that  $\sum_{i=1}^n \text{Reg}_i^T \geq 0 \Rightarrow \sum_{i=1}^n \sum_{t=1}^T \|x_i^t - x_i^{t-1}\|^2 = O(1)$ . Optimistic learning

**Key observation:**  $\text{NEGAP}(x^t) \leq O(1) \sum_{i=1}^n \|x_i^t - x_i^{t-1}\|$ . Holds for optimistic gradient descent

$$\sum_{t=1}^T (\text{NEGAP}(x^t))^2 \leq O(1) \implies \text{NEGAP}(x^{t^*}) = O\left(\frac{1}{\sqrt{T}}\right)$$

There is a matching lower bound

# Potential games

- So far we have focused on **strictly competitive** games
- What about **cooperative games**?
- E.g., **identical interest** games or **potential games**  $u_i = \nabla \Phi$

**Theorem.** (*Optimistic*) *gradient descent converges to  $\epsilon$ -Nash equilibria after  $O(1/\epsilon^2)$  iterations in potential games.*

**Corollary.** *Optimistic gradient descent guarantees  $\text{Reg}_i^T = O(1)$ .*

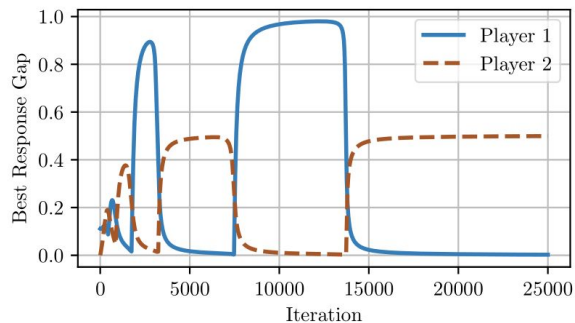
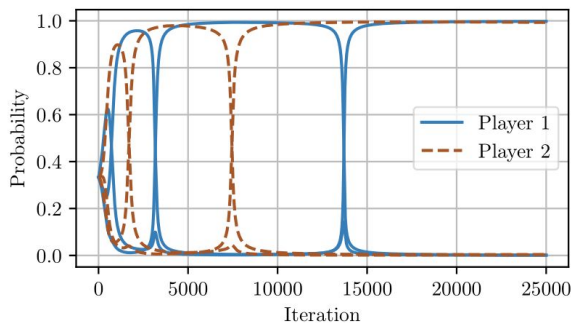
# General games?

- We have focused on **restricted** classes of games (e.g. 2p0s, potential)
- Can we hope to extend **last-iterate convergence** to general games?
- No! **Last-iterate convergence** is inherently tied to **Nash equilibria**
- And **Nash equilibria** are hard to compute
- A single iterate is **uncorrelated** (correlation requires multiple iterates)
- Lack of convergence is **inherent** in no-regret learning in games



# Small variation despite cycling

- We saw earlier that  $\sum_{t=1}^T \sum_{i=1}^n \|x_i^t - x_i^{t-1}\|^2 = O(\log T)$
- The dynamics are **still cycling** (in general)
- **Small variation** does not always imply Nash equilibria



# Social welfare of no-regret dynamics

- No-regret dynamics converge to *some* **equilibrium**
- Is that enough?
- Some **equilibria** are better than others; e.g., in **(social) welfare**
- Converge to **equilibria** with **high welfare**
- What is the **welfare** of no-regret dynamics?
- Maximizing welfare of coarse correlated equilibria is **NP-hard**
- Content with **near-optimal welfare** for broad classes of games

# Smooth games

**Definition** (Roughgarden 2015, JACM). A game is  $(\lambda, \mu)$ -**smooth** if there exists  $x^*$  s.t.

$$\sum_{i=1}^n u_i(x_i^*, x_{-i}) \geq \lambda \text{OPT} - \mu \sum_{i=1}^n u_i(x), \forall x.$$

- **Robust price of anarchy:**  $\text{rPoA} = \frac{\lambda}{1 + \mu}$
- $\text{rPoA} = 0.5$  in **simultaneous second-price auctions**
- $\text{rPoA} = 0.4$  in **congestion games with linear latency functions**

# Regret minimization in smooth games

**Theorem** (Roughgarden 2015, JACM). *In any  $(\lambda, \mu)$ -smooth game, any no-regret dynamics attain a  $\text{rPoA}$  fraction of the *optimal welfare*.*

Proof:

$$\sum_{i=1}^n \text{Reg}_i^T = \sum_{i=1}^n \sum_{t=1}^T (u_i(x_i^*, x_{-i}^t) - u_i(x^t)) \geq \lambda T \text{OPT} - (1 + \mu) \sum_{t=1}^T \sum_{i=1}^n u_i(x^t).$$

$$\frac{1}{T} \sum_{t=1}^T \text{SW}(x^t) \geq \frac{\lambda}{1 + \mu} \text{OPT} - \frac{1}{1 + \mu} \frac{1}{T} \sum_{i=1}^n \text{Reg}_i^T.$$

# Improved welfare using optimism

**Theorem** (A-Panageas-Farina-Sandholm 2022, ICML). *Optimistic learning dynamics have the following property:*

- Either they converge to an  $O(\epsilon)$ -Nash equilibrium;
- Or the welfare **outperforms the robust price of anarchy**:

$$\frac{1}{T} \sum_{t=1}^T \text{SW}(x^t) \geq \frac{\lambda}{1 + \mu} \text{OPT} + \epsilon^2.$$

# Takeaways

- **Cycling** can lead to **improved welfare** guarantees
- The *further away* from Nash, the *larger* the **improvement in welfare**
- Interesting interplay between **regret**, **convergence**, and **welfare**
- Many interesting open problems!

# References

- Roughgarden 2015: Intrinsic Robustness of the Price of Anarchy, JACM
- Wei-Lee-Zhang-Luo 2021: Linear Last-Iterate Convergence in Constrained Saddle-Point Optimization, ICLR
- Anagnostides-Panageas-Farina-Sandholm 2022: On Last-Iterate Convergence Beyond Zero-Sum Games, ICML
- Tseng 1995: On Linear Convergence of Iterative Methods for the Variational Inequality Problem, J. Comp. & Appl. Math.

# References

- Gilpin-Peña-Sandholm 2012: First-order algorithm with  $O(\ln(1/\epsilon))$  convergence for  $\epsilon$ -equilibrium in two-person zero-sum games, Math. Prog.
- Chen-Peng 2020: Hedging in games: Faster convergence of external and swap regrets, NeurIPS
- Syrgkanis-Agarwal-Luo-Schapire 2015: Fast convergence of regularized learning in games, NIPS
- Anagnostides-Farina-Luo-Lee-Kroer-Sandholm 2022: Uncoupled Learning Dynamics with  $O(\log T)$  Swap Regret, NeurIPS
- Farina-Grand-Clément-Kroer-Lee-Luo 2023: Regret Matching+:(In) Stability and Fast Convergence in Games