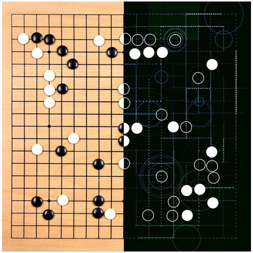
Beyond Equilibrium Learning

Chi Jin

Princeton University.

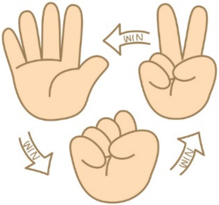
Problems of Interests



Games / strategic decision making against one or more adaptive opponents.

Normal-Form Games (NFGs)

Represent games as matrices (tensors):



P2 / P1	Rock	Paper	Scissors
Rock	(0, 0)	(-1, 1)	(1, -1)
Paper	(1, -1)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

In general, specify utility $u_i(a_1, \ldots, a_n)$ for $i \in [n]$.

Sequential games can be represented as big NFGs, where
actions in NFGs $\Leftrightarrow$ policies in sequential games.

Overview

1. **Standard game theory**:
   — equilibrium and learning algorithms

2. **Beyond equilibrium learning I**:
   — rationalizability

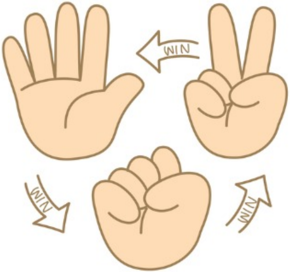
3. **Beyond equilibrium learning II**:
   — symmetry and equal share

4. **Conclusion**

Standard Game Theory
— equilibrium and learning algorithms

Optimal Strategy

What is the optimal strategy?



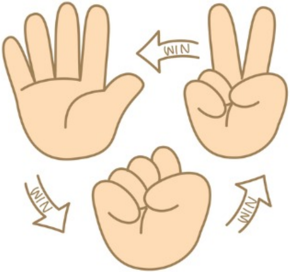
The strategies of games may not have a linear relation (i.e. cyclic structure)

strategy A > strategy B > strategy C > strategy A

Nash Equilibrium

A (mixed) strategy π_i for the i^{th} player is a probability over action set $\mathcal{A}_i$.

Nash equilibrium (NE): a *product* policy $\pi = \pi_1 \times \cdots \times \pi_m$, where no player can gain by deviating from her own policy while fixing other players' policies.



E.g., in rock-paper-scissor, an NE is $\pi_1 = \pi_2 = \text{Uniform}(\mathcal{A})$.

Correlated Strategy

Nash equilibrium assumes each agents play independently.

In general-sum games, we may prefer correlated strategies (win-win):

P1 \ P2	R	P	S
R	(1, 1)	(-1, 1)	(1, -1)
P	(1, -1)	(1, 1)	(-1, 1)
S	(-1, 1)	(1, -1)	(1, 1)

Table 1: Modified rock-paper-scissor

P1 \ P2	R	P	S
R	$1/3$	0	0
P	0	$1/3$	0
S	0	0	$1/3$

Table 2: A correlated strategy that can be realized by **shared random bits**

Correlated Equilibrium

Correlated equilibrium (CE): a **correlated** policy π, where no player can gain by deviating her own policy while fixing other players' policies, **if the deviator can still see the shared random bits from the correlated policy**.

Coarse correlated equilibirum (CCE): a correlated policy π, where ..., if the deviator can no longer see the shared random bits.

Table 2 is a CCE but not a CE. In general, NE $\subset$ CE $\subset$ CCE.

No-regret Learning

Originally for adversarial bandits. A powerful tool for learning equilibrium.

Each round: player chooses a mixed strategy μ_t, environment chooses an adversarial loss ℓ_t. Regret is measured against the best action in hindsight.

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} \langle \mu_t, \ell_t \rangle - \min_{a \in \mathcal{A}} \sum_{t=1}^{T} \langle a, \ell_t \rangle \leq o(T).$$

From No-regret to Learning Equilibrium

Hedge algorithm: performs exponential weight updates:

$$\mu^{t+1}(a) \propto \mu^{t}(a) e^{-\eta_t \ell_t(a)}, \quad \text{for} \quad \forall a \in \mathcal{A}.$$

where η_t is the learning rate. Hedge achieves $\tilde{O}(\sqrt{T})$ regret.

All players run no-regret algorithms independently, the average policy

$$\frac{1}{T} \sum_{t=1}^{T} \mu_1^t \times \cdots \times \mu_n^t \to \text{CCE}$$

2p0s: marginalize CCE $\to$ NE; **General**: finding NE is PPAD-hard.

Beyond Equilibrium Learning I:
— rationalizability

Collaborators

Yuanhao Wang
Princeton

Dingwen Kong
MIT

Yu Bai
Salesforce

Rationalizability

A rational agent should not play dominated actions, which is strictly worse than another strategy no matter what opponent plays

P1 \ P2	B1	B2
A1	(1, 3)	(2, 2)
A2	(2, 2)	(1, 1)

Iterative Dominance Elimination

Eliminating dominated actions iteratively

P1 \ P2	B1	B2
A1	(1, 3)	(2, 2)
A2	(2, 2)	(1, 1)

P1 \ P2	B1
A1	(1, 3)
A2	(2, 2)

define rationalizable $\Leftrightarrow$ play iteratively un-dominated actions.

An action is Δ-rationalizable if it remains after iteratively eliminating Δ-dominated actions.

Rationalizability vs Equilibrium

- NEs and CEs are rationalizable.
- ϵ-CE can be entirely supported on iteratively dominated actions, unless $\epsilon = \mathcal{O}(2^{-A})$ [Wu et al., 2021]
- CCEs are not necessarily rationalizable [Viossat & Zapechelnyuk 2013].

P1 \ P2	A	B	C
A	(2, 2)	(1, 1)	(-4, -4)
B	(1, 1)	(0, 0)	(-1, -1)
C	(-4, -4)	(-1, -1)	(-2, -2)

Main Question

Can we efficiently learn equilibria that are also rationalizable?

Bandit feedback: not knowing game rule; at each round, player i only observes a random payoff $U_i(a_1, \ldots, a_n)$.

A Naive Approach

A direct two-stage approach to learn rationalizable equilibria:

1. **identify** the set of all Δ-rationalizable actions
2. learn equilibria in the subgame restricted to these rationalizable actions

This incurs $\Omega(A^n)$ sample complexity, where n is the number of players.

Verifying action dominance requires the enumeration of the joint action space of other players, which is exponentially large.

Our Algorithm

Rationalizable Hedge

Find a rationalizable action profile and initialize joint policy $\{\theta_i^{(0)}\}_{i=1}^n$ there.

for $t = 1, \ldots, T$,

estimate loss by $\ell_i^{(t)}(a)$ by playing $\theta^{(t)}$ for M times.

perform Hedge update $\mu_i^{t+1}(a) \propto \mu_i^t(a) e^{-\eta_t \ell_i^{(t)}(a)}$

output: policy μ_i^T after eliminating small probability actions & renormalizing.

- Identifying whether one action is rationalizable is hard, but finding one rationalizable action is not hard.
- Our algorithm guarantees that the policy $\{\mu_i^t\}$ is always mostly supported on rationalizable actions across all iterates.

Theoretical Guarantees

Theorem [Wang, Kong, Bai, Jin, 2023]

Rationalizable Hedge finds Δ-rationalizable ϵ-CCEs within sample complexity:

$$\tilde{\mathcal{O}}\left(\frac{LNA}{\Delta^2} + \frac{NA}{\epsilon^2}\right)$$

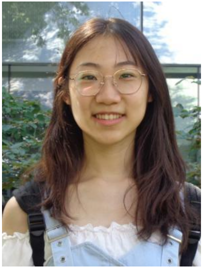
$L < NA$ is the minimum elimination length.

First polynomial sample algorithms for learning rationalizable equilibria!

We extend it to find rationalizable CEs by no-swap-regret.

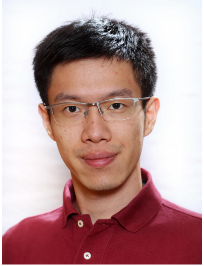
Beyond Equilibrium Learning II:
— symmetry and equal share

Collaborators



Jiawei Ge
Princeton

Yuanhao Wang
Princeton



Wenzhe Li
Princeton

Symmetric Games and Equal Share

Many games are designed to be fair and symmetric to all players:

Suppose the game is constant-sum with total payoff C per game.

Baseline: achieve at least equal share! i.e., payoff C/n per game.

Focus on symmetric zero-sum games.

Equal Share vs Equilibrium

Two-player zero-sum games:

- Nash is *unique.
- A Nash strategy is non-exploitable — achieve at least 0 payoff (equal share) no matter what the opponent plays!

Multi-player ($n > 2$) zero-sum games:

- Nash/CE/CCEs are all non-unique.
- Nash does not guarantee equal share (even against fixed opponents)!

Consider three-player majority vote games: action set $\{0, 1\}$, majority gets 1 while minority gets -2. Both $(1, 1, 1)$ and $(0, 0, 0)$ are Nash.

Prior Algorithms

Self-play

for $t = 1, \ldots, T$,

all agents play policy θ_t.

the main agent perform updates to obtain new policy θ_{t+1}.

- main ingradient for SOTA systems for Poker, Mahjoon, Diplomacy, etc.
- can use gradient updates, Hedge updates, ...

Claim: self-play from scratch does not guarantee equal share.

Again, consider three-player majority vote games.

Main Questions

What is the right solution concept to achieve equal share?

Can we design provably efficient algorithms for achieving equal share?

Solution Concept I

Observation 1: If opponents are permitted to adopt different strategies, there are games agent can't obtain equal share no matter what she does.

Consider three-player minority vote games: action set $\{0, 1\}$, majority gets -2 while minority gets -1. Two opponents play 0 and 1 separately.

Takeaway: Must consider settings opponents deploy identical strategies, — not bad in games with a large player base.

$$\max_{x_1} \min_{x_2,\cdots,x_n} U_1(x_1, \cdots, x_n) \leq \min_{x_2,\cdots,x_n} \max_{x_1} U_1(x_1, \cdots, x_n) \leq \min_{x} \max_{x_1} U_1(x_1, x^{\otimes n-1})$$

Solution Concept II

Observation 2: There are games where no non-exploitable strategies do not exist even after restricting all opponents to play identical strategy.

Takeaway: to achieve equal share, the agent has to model opponents.

$$\max_{x_1} \min_{x} U_1(x_1, x^{\otimes n-1}) \leq \min_{x} \max_{x_1} U_1(x_1, x^{\otimes n-1}) = 0,$$

Efficient Algorithms

Stationary opponents (with identical strategies):

- The best response can achieve equal share.
- Run no-regret algorithms.

Adaptive (but slowly changing) opponents:

- Run no-dynamic-regret algorithms.

Experiments

We can construct

- symmetric zero-sum games
- stationary and identical policies for opponents

that robustly breaks all meta-algorithms in prior SOTA systems.

SDG	SP_scratch / SP_BC / SP_BC_reg	BR_BC
Utility	-12.67	**1.00**
Exploitability	-29.00	-29.00

Summary

Standard game theory:

- Nash, CE, CCEs
- no-regret learning algorithms

Rationalizability

- limitation of applying standard game theory
- Rationalizable Hedge

Equal Share in Symmetric Games

- identify the right solution concepts
- develop efficient algorithms