Is Learning in Games Good for the Learners?

William Brown (Columbia)
Jon Schneider (Google)
Kiran Vodrahalli (Google)

Is Learning in Games Good for the Learners?

Setting:

- 2-player general-sum games $G = (A, B)$, played for T rounds.

Many prior works asking:

- How fast do learning algorithms converge to (coarse) correlated equilibria?
- How do (coarse) correlated equilibria compare to optimal welfare for specific classes of games?

Questions we address:

- Is it actually *good* for agents (in terms of welfare) to run a no-(swap)-regret algorithm against a no-(swap)-regret opponent?
- How does the answer depend on the details of the opponent's algorithm?
- How does the answer depend on structural properties of the game?
- How does the strategy change if the game is initially known vs. unknown?

Generalized (Φ_A, Φ_B)-Equilibria

We consider "generalized equilibria" with asymmetric regret constraints Φ_A and Φ_B for players A and B.

- Focus: "linear" constraints Φ, which includes internal (I), external (E), and unconstrained $(\emptyset)$
- Generalizes CE, CCE, etc.

Motivation:

- Each pair of regret constraints (Φ_A, Φ_B) for a game G corresponds to a polytope;
- For any fixed game G, we can compute upper and lower utility bounds for each player, knowing only their regret constraints.

Theorem 1:

For any (Φ_A, Φ_B)-equilibrium Ψ in a game G, there exists a pair of algorithms $(\mathscr{L}_A, \mathscr{L}_B)$ such that:

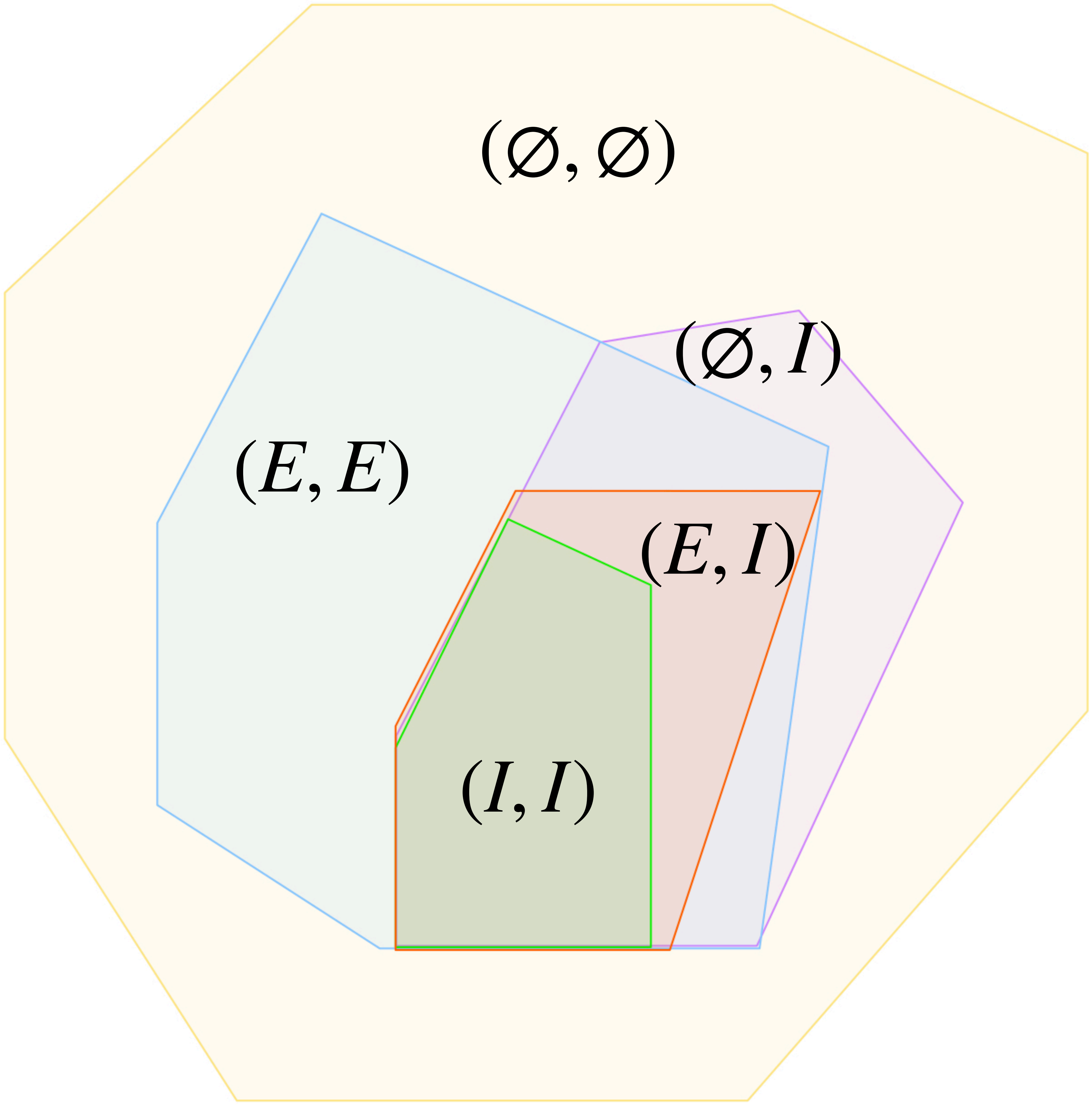
- $\mathscr{L}_A$ and $\mathscr{L}_B$ converge to Ψ when played together;
- $\mathscr{L}_A$ and $\mathscr{L}_B$ are no-Φ_A-regret and no-Φ_B-regret, respectively, against arbitrary adversaries.

We use this result to analyze reward-regret tradeoffs by comparing best-case/worst-case utility for a player under different regret pairs.

Generalized (Φ_A, Φ_B)-Equilibria

Example sets of generalized equilibria:

- All (coarse) correlated equilibria
- All joint strategy profile distributions
- All possible convergent profile distributions against a no-(internal)-regret learner



Results via Generalized Equilibria

Some of our results:

- The optimal $(\varnothing, I)$-equilibrium for Player A matches the Stackelberg value of a game, which is attainable against a no-internal/no-swap learner;
- We tightly characterize when some (+ all) pairs of no-swap algorithms form a Nash equilibrium for the "metagame" (where players choose algorithms);
- In "almost all" games without a pure Nash equilibrium (w.r.t. measure), the Stackelberg value beats the best correlated equilibrium;
- There is an LP which characterizes the best reward attainable against "mean-based" learners, which can be worse than the best $(\varnothing, E)$-equilibrium

Takeaways:

- The Stackelberg value is always attainable against a no-Φ-regret learner (by playing the Stackelberg strategy);
- The Stackelberg value is often optimal and strictly better than all (coarse) correlated equilibria, and can only be improved if more is known about the opponent's algorithm.

Learning Stackelberg with a No-Regret Opponent

The Stackelberg strategy is easy to compute and implement if the game is known. But what if we don't know our opponent's reward function?

- We give reductions from "best response query" offline algorithms to adaptive strategies against no-regret opponents
- Key idea: if we play a mixed strategy for long enough, a no-regret opponent will eventually best-respond

Theorem:

If the Stackelberg equilibrium Ψ for a game G is learnable with Q best-response queries, then:

- Ψ is learnable in exp(Q) rounds against any no-regret learner
- Ψ is learnable in poly(Q) rounds against any dynamic/adaptive-regret learner
- There are "mean-based" learners where exp(Q) are required to learn Ψ