Heads-up limit hold’em poker is solved

Main Presentation

Counterfactual Regret Minimization (CFR)

- Iterative algorithm that converges to Nash equilibrium.
- At each iteration the current strategy plays at least one game against itself.
- The strategy is updated at reached information sets.
- The new mixture at each information set is chosen by **regret matching**.

Regret Matching

Mix in proportion to counterfactual regret:

$$P(a) = \frac{R(a)}{\sum_{R(a')>0} R(a')}$$

Actions with *R ≤ 0* are unplayed unless all actions have *R ≤ 0*, in which case regret matching mixes uniformly.

Counterfactual Regret

For a strategy profile σ, an action a and an information set I at time T:

$$R(\sigma, a, I, T) = \sum_{t=1}^{T} Eu(\sigma^t \rightarrow a|I) - Eu(\sigma^t|I)$$

where $\sigma^t \rightarrow a = \sigma^t$ everywhere except that a is played deterministically at I.

Properties of CFR

- A strategy’s regret is bounded by the sum of its counterfactual regrets at every information set.

- Each information set’s average counterfactual regret declines over time.

CFR$^+$

Negative counterfactual regrets are reset to 0 each iteration:

$$\begin{aligned} R(\sigma, a, I, T) = \max(&R(\sigma, a, I, T-1), 0) \\ &+ Eu(\sigma^T \to a|I) - Eu(\sigma^T|I) \end{aligned}$$

So as soon as a strategy would do well, it gets played.

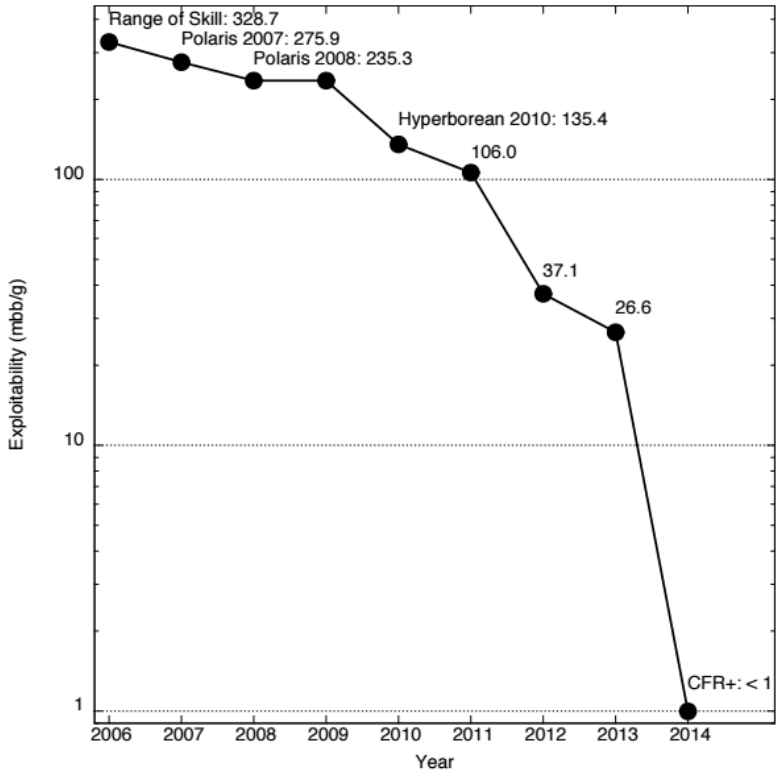
Implementation improvements

- Compression
  - Use fixed-point arithmetic.
  - Sort the values for before zipping them.

- Parallelization
  - 110,565 subgames were split across 199 workers.
  - 1 master responsible for the top of the game tree.

Exploitability

- How much the strategy loses if the opponent best-responds.
- Bounds the value of the game.
- Can be computed exactly as discussed last time.



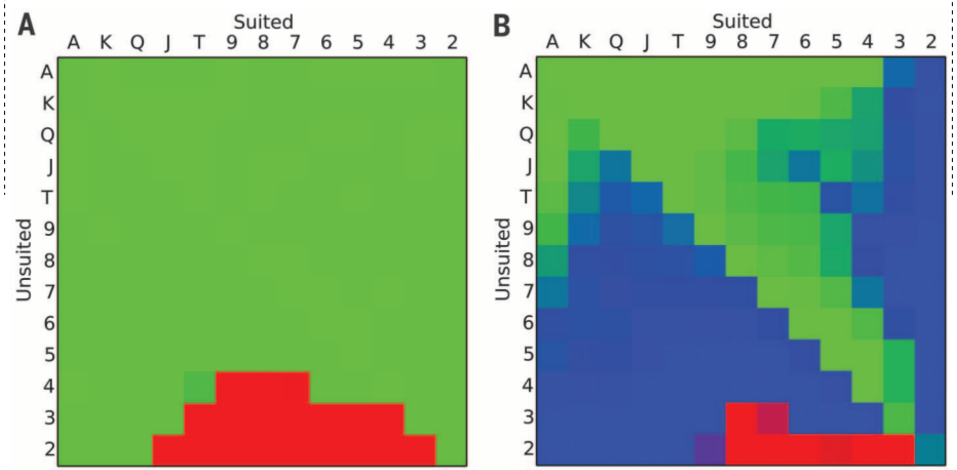
Essentially-weakly solved

- Exploitability below 1mbb per game.

- With such low exploitability, there is a 5% chance of beating a best-responding opponent over a lifetime of games.

What we learned about poker

- The dealer has an advantage of ~0.1 big blinds per hand.
- In this equilibrium, limping and capping in the pre-flop betting are rare.
- Other equilibria are possible.
    - They must have the same value.
    - They could have different strategies.



Broader applicability

- Large zero-sum incomplete information games can be solved.

- Discussion question: what else can we model now that we have CFR+?