Some Approaches for Large Imperfect-Information Games

MIT 6.S890

Samuel Sokota; October 29, 2024

Question:

How can we approach large imperfect-information games?

Question:

How can we approach large imperfect-information games?

- Tabular game-theory?

Question:

How can we approach large imperfect-information games?

- Tabular game-theory?
- Deep reinforcement learning?

Question:

How can we approach large imperfect-information games?

- Tabular game-theory?
- Deep reinforcement learning?
- “Deepified” game-theory?

Question:

How can we approach large imperfect-information games?

- Tabular game-theory?
- Deep reinforcement learning?
- “Deepified” game-theory?
- “Game-theorified” deep reinforcement learning?

Question:

How can we approach large imperfect-information games?

- CFR?

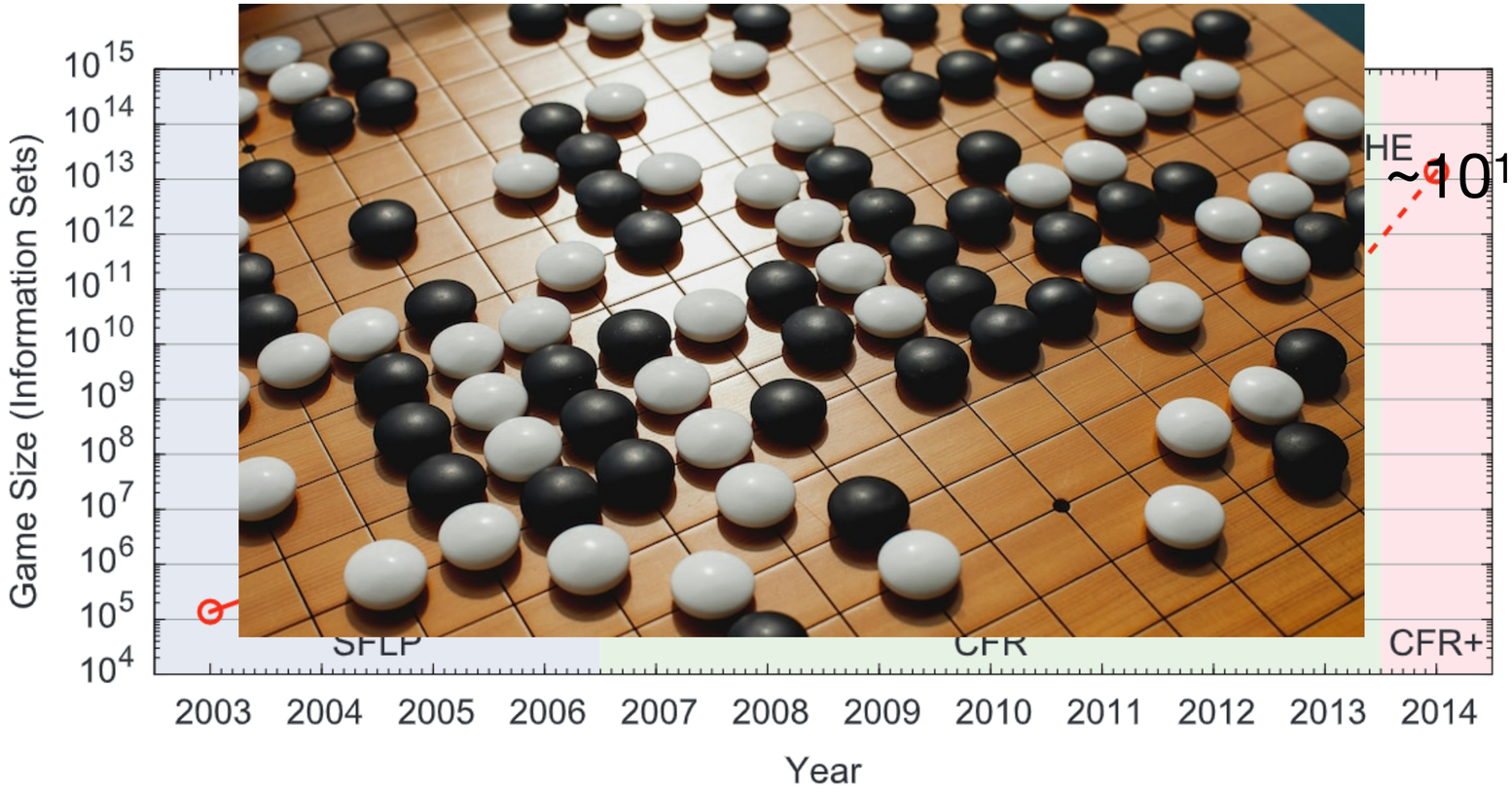


Fig. 2. Increasing sizes of imperfect-information games solved over time measured in unique information sets (i.e., after symmetries are removed). The shaded regions refer to the technique used to achieve the result; the dashed line shows the result established in this paper.

Question:

How can we approach large imperfect-information games?

- Deep reinforcement learning?

Question:

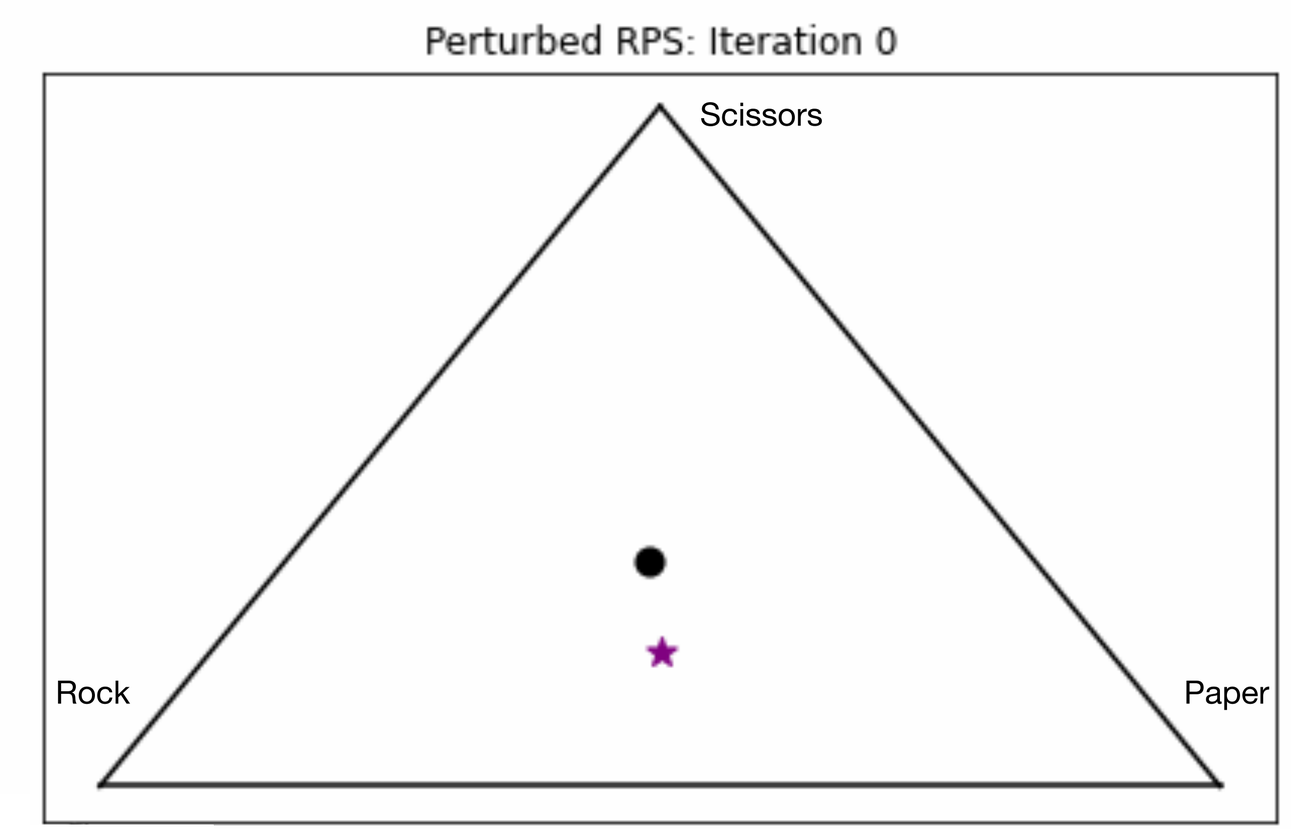
How can we approach large imperfect-information games?

- Deep reinforcement learning?

Question:

How can we approach large imperfect-information games?

- Deep reinforcement learning?



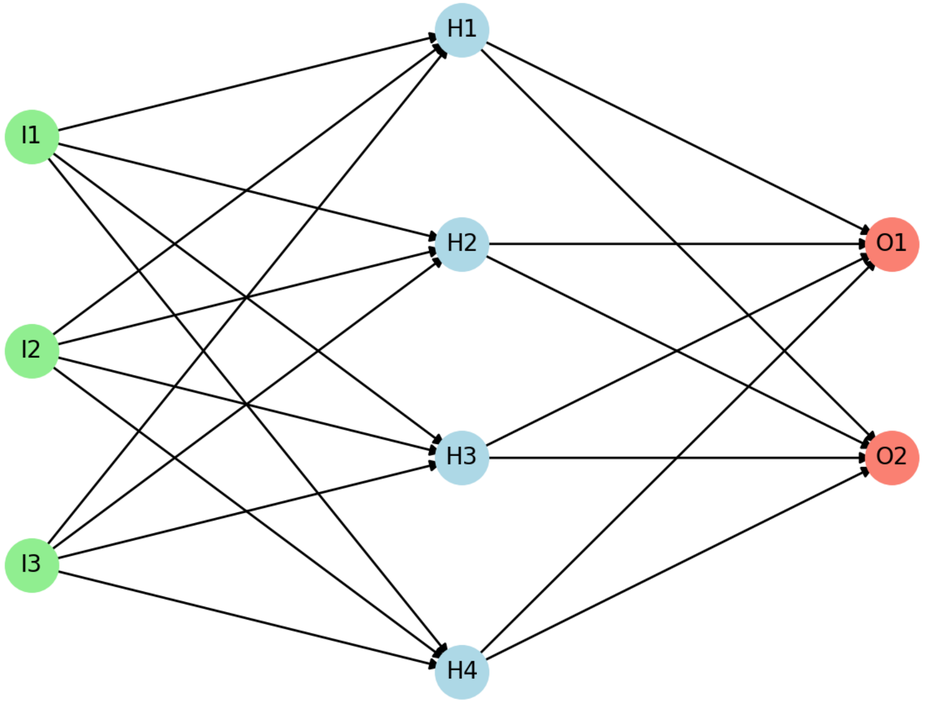
Question:

How can we approach large imperfect-information games?

- “Deepified” game-theory?

Sub-question:

How can we “deepify” game-theoretic approaches?



Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow \text{uniform\_policy}()$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
    $\pi \leftarrow$ best_response($\bar{\pi}$)

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
    $\pi \leftarrow$ best_response($\bar{\pi}$)
    $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

$\Pi \leftarrow \{$uniform_policy()$\}$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

$\Pi \leftarrow \{$uniform_policy()$\}$
while is_time_left() **do**
  $\pi_* \leftarrow$ nash(Π)

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

$\Pi \leftarrow \{$uniform_policy()$\}$
while is_time_left() **do**
  $\pi_* \leftarrow$ nash(Π)
  $\pi \leftarrow$ best_response(π_*)

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

$\Pi \leftarrow \{\text{uniform\_policy}()\}$
while is_time_left() **do**
  $\pi_* \leftarrow \text{nash}(\Pi)$
  $\pi \leftarrow \text{best\_response}(\pi_*)$
  $\Pi \leftarrow \Pi \cup \{\pi\}$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

$\Pi \leftarrow \{$uniform_policy()$\}$
while is_time_left() **do**
  $\pi_* \leftarrow \text{nash}(\Pi)$
  $\pi \leftarrow$ best_response(π_*)
  $\Pi \leftarrow \Pi \cup \{\pi\}$
end while
return π_*

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 1: Use deep reinforcement learning to approximate best response

Algorithm 1 Fictitious Play

$\bar{\pi} \leftarrow$ uniform_policy()
while is_time_left() **do**
  $\pi \leftarrow$ best_response($\bar{\pi}$)
  $\bar{\pi} \leftarrow$ update_average($\bar{\pi}, \pi$)
end while
return $\bar{\pi}$

Algorithm 2 Double Oracle

$\Pi \leftarrow \{$uniform_policy()$\}$
while is_time_left() **do**
  $\pi_* \leftarrow$ nash(Π)
  $\pi \leftarrow$ best_response(π_*)
  $\Pi \leftarrow \Pi \cup \{\pi\}$
end while
return π_*

Sub-question:

How can we “deepify” game-theoretic approaches?

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 2: Use deep learning to approximate regret

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 2: Use deep learning to approximate regret

$$\sigma^{t+1}(I,a) = \frac{R_+^t(I,a)}{\sum_{a' \in A(I)} R_+^t(I,a')}$$

Sub-question:

How can we “deepify” game-theoretic approaches?

- Idea 2: Use deep learning to approximate regret

$$\sigma^{t+1}(I,a) = \frac{R_+^t(I,a)}{\sum_{a' \in A(I)} R_+^t(I,a')}$$

“Deepfied” Game-Theoretic Approaches

Deep Reinforcement Learning from Self-Play in Imperfect-Information Games

Johannes Heinrich
University College London, UK
j.heinrich@cs.ucl.ac.uk

David Silver
University College London, UK
d.silver@cs.ucl.ac.uk

Fictitious Play + DRL (NFSP)

“Deepfied” Game-Theoretic Approaches

Deep Reinforcement Learning from Self-Play in Imperfect-Information Games

Johannes Heinrich
University College London, UK
j.heinrich@cs.ucl.ac.uk

David Silver
University College London, UK
d.silver@cs.ucl.ac.uk

Fictitious Play + DRL (NFSP)

A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning

Marc Lanctot
DeepMind
lanctot@

Vinicius Zambaldi
DeepMind
vzambaldi@

Audrūnas Gruslys
DeepMind
audrunas@

Angeliki Lazaridou
DeepMind
angeliki@

Karl Tuyls
DeepMind
karltuyls@

Julien Pérolat
DeepMind
perolat@

David Silver
DeepMind
davidsilver@

Thore Graepel
DeepMind
thore@

Double Oracle + DRL (PSRO)

“Deepfied” Game-Theoretic Approaches

Deep Reinforcement Learning from Self-Play in Imperfect-Information Games

Johannes Heinrich
University College London, UK
j.heinrich@cs.ucl.ac.uk

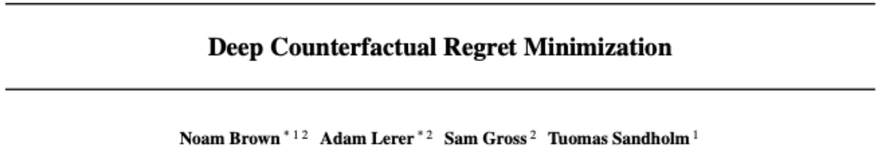
David Silver
University College London, UK
d.silver@cs.ucl.ac.uk

Fictitious Play + DRL (NFSP)

A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning

Marc Lanctot DeepMind lanctot@
Vinicius Zambaldi DeepMind vzambaldi@
Audrūnas Gruslys DeepMind audrunas@
Angeliki Lazaridou DeepMind angeliki@
Karl Tuyls DeepMind karltuyls@
Julien Pérolat DeepMind perolat@
David Silver DeepMind davidsilver@
Thore Graepel DeepMind thore@

Double Oracle + DRL (PSRO)



Deep Counterfactual Regret Minimization

Noam Brown [*1 2] **Adam Lerer** [*2] **Sam Gross** [2] **Tuomas Sandholm** [1]

CFR + DL (Deep CFR)

“Deepfied” Game-Theoretic Approaches

Pros and Cons

“Deepfied” Game-Theoretic Approaches

Pros and Cons

+ Clear theoretical foundation

“Deepfied” Game-Theoretic Approaches

Pros and Cons

+ Clear theoretical foundation

- Approximate best responses are expensive

“Deepfied” Game-Theoretic Approaches

Pros and Cons

+ Clear theoretical foundation

- Approximate best responses are expensive

- Fictitious play & double oracle can converge slowly

“Deepfied” Game-Theoretic Approaches

Pros and Cons

+ Clear theoretical foundation

- Approximate best responses are expensive

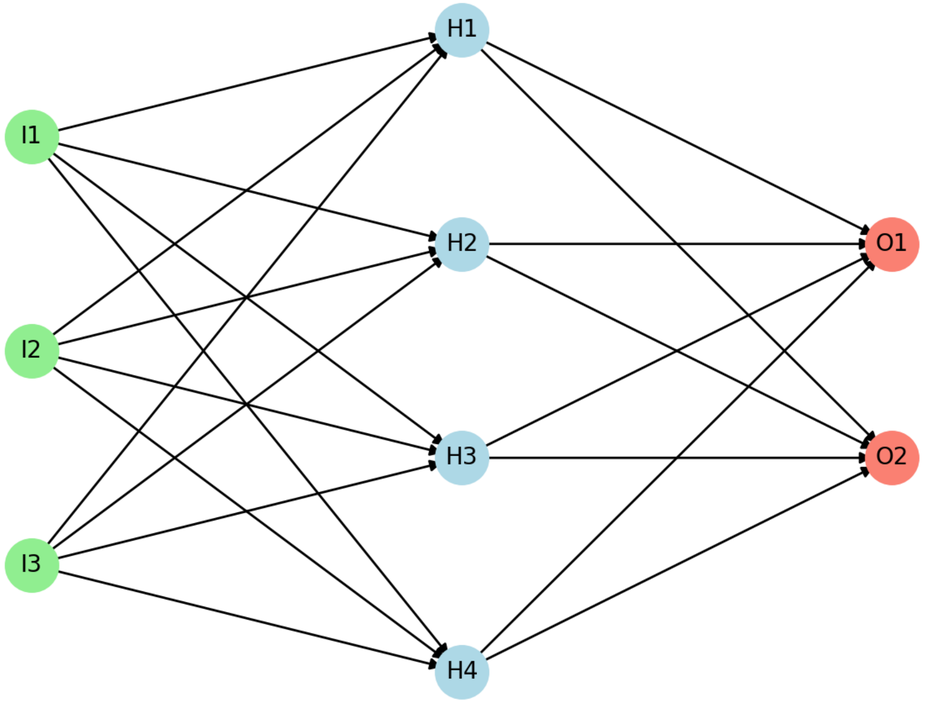
- Fictitious play & double oracle can converge slowly

- Importance sampling can cause high variance

What about the other direction?

What about the other direction?

Can we “game-theorify” deep reinforcement learning?



What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size
3. Regularize policy

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size

} Online mirror descent

3. Regularize policy

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size

} Online mirror descent

3. Regularize policy

$$\pi_{t+1} = \arg\max_{\pi} \langle q, \pi \rangle - \frac{1}{\eta} \mathrm{KL}(\pi, \pi_t)$$

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size

} Online mirror descent

3. Regularize policy

$$\pi_{t+1} = \arg\max_{\pi} \langle q, \pi \rangle - \frac{1}{\eta} \mathrm{KL}(\pi, \pi_t) - \alpha \mathrm{KL}(\pi, \rho)$$

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size

} Online mirror descent

3. Regularize policy

“Magnetic” mirror descent

$$\pi_{t+1} = \arg\max_{\pi} \langle q, \pi \rangle - \frac{1}{\eta}\mathrm{KL}(\pi, \pi_t) - \alpha \mathrm{KL}(\pi, \rho)$$

A Unified Approach to Reinforcement Learning, Quantal Response Equilibria, and Two-Player Zero-Sum Games

Samuel Sokota* Carnegie Mellon University ssokota@andrew.cmu.edu

Ryan D’Orazio* Mila, Université de Montréal ryan.dorazio@mila.quebec

J. Zico Kolter Carnegie Mellon University zkolter@cs.cmu.edu

Nicolas Loizou Johns Hopkins University nloizou@jhu.edu

Marc Lanctot DeepMind lanctot@deepmind.com

Ioannis Mitliagkas Mila, Université de Montréal ioannis@mila.quebec

Noam Brown Meta AI noambrown@meta.com

Christian Kroer Columbia University ck2945@columbia.edu

What about the other direction?

Can we “game-theorify” deep reinforcement learning?

Modern deep policy gradient algorithms:

1. Maximize value
2. Control update size

} Online mirror descent

3. Regularize policy

“Magnetic” mirror descent

$$\pi_{t+1} = \arg\max_{\pi} \langle q, \pi \rangle - \frac{1}{\eta}\mathrm{KL}(\pi, \pi_t) - \alpha \mathrm{KL}(\pi, \rho)$$

Magnet

A Unified Approach to Reinforcement Learning, Quantal Response Equilibria, and Two-Player Zero-Sum Games

Samuel Sokota* Carnegie Mellon University ssokota@andrew.cmu.edu

Ryan D’Orazio* Mila, Université de Montréal ryan.dorazio@mila.quebec

J. Zico Kolter Carnegie Mellon University zkolter@cs.cmu.edu

Nicolas Loizou Johns Hopkins University nloizou@jhu.edu

Marc Lanctot DeepMind lanctot@deepmind.com

Ioannis Mitliagkas Mila, Université de Montréal ioannis@mila.quebec

Noam Brown Meta AI noambrown@meta.com

Christian Kroer Columbia University ck2945@columbia.edu

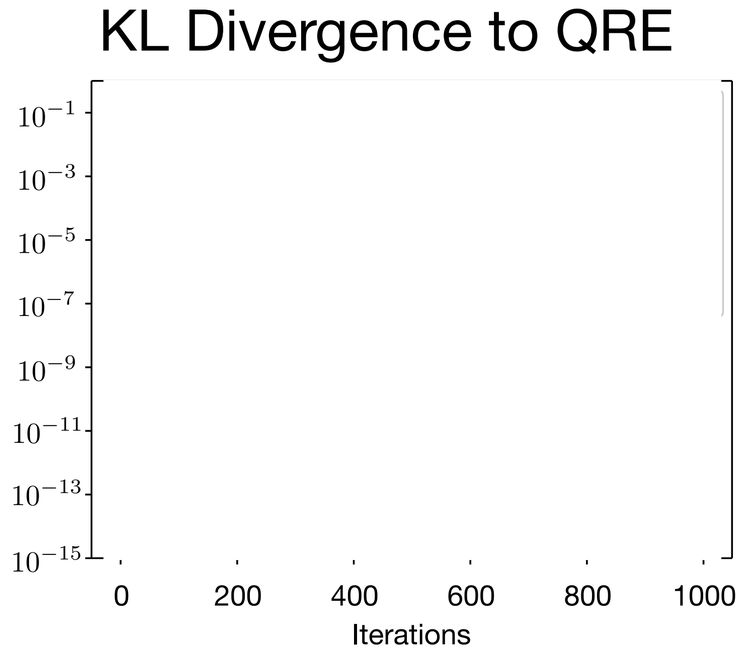
What can we say about magnetic mirror descent?

What can we say about magnetic mirror descent?

In two-player zero-sum normal-form games, if $\eta \leq \alpha / L^2$ magnetic mirror descent converges exponentially fast to a regularized equilibrium in self-play.

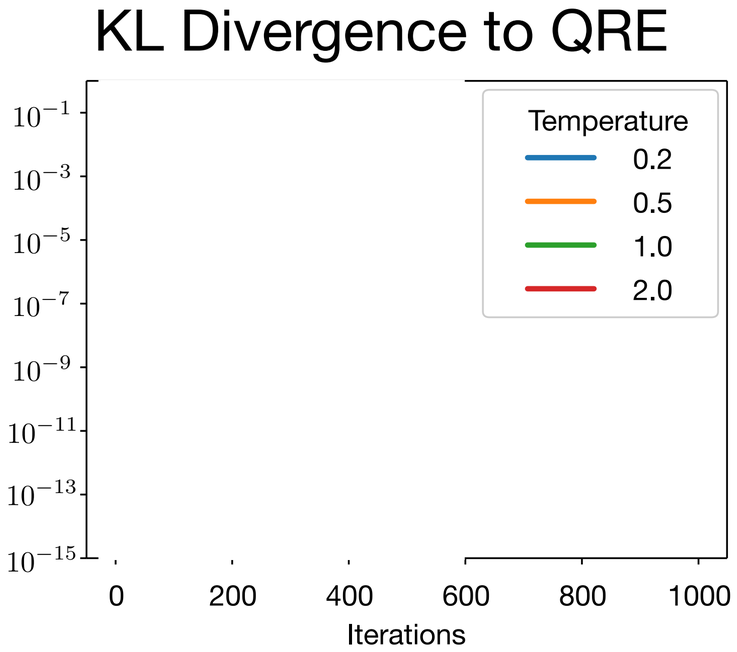
What can we say about magnetic mirror descent?

In two-player zero-sum normal-form games, if $\eta \leq \alpha / L^2$ magnetic mirror descent converges exponentially fast to a regularized equilibrium in self-play.



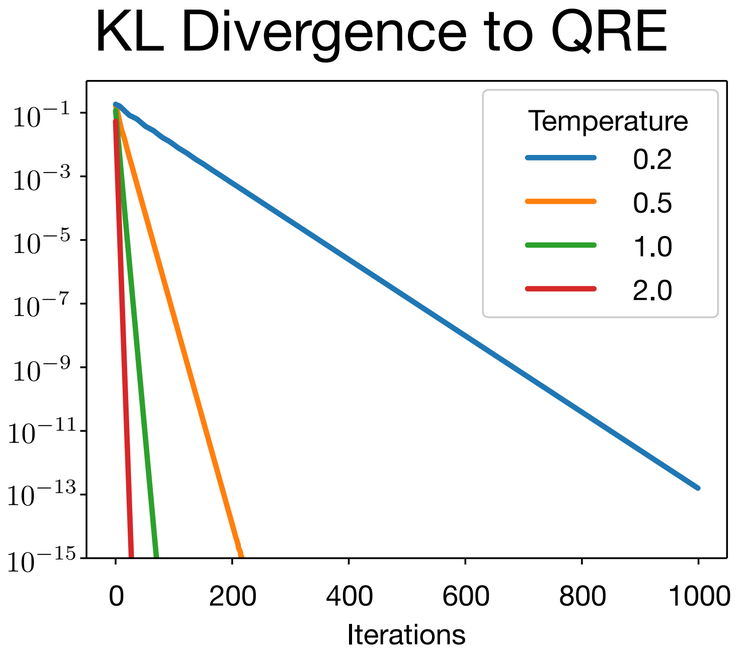
What can we say about magnetic mirror descent?

In two-player zero-sum normal-form games, if $\eta \leq \alpha / L^2$ magnetic mirror descent converges exponentially fast to a regularized equilibrium in self-play.



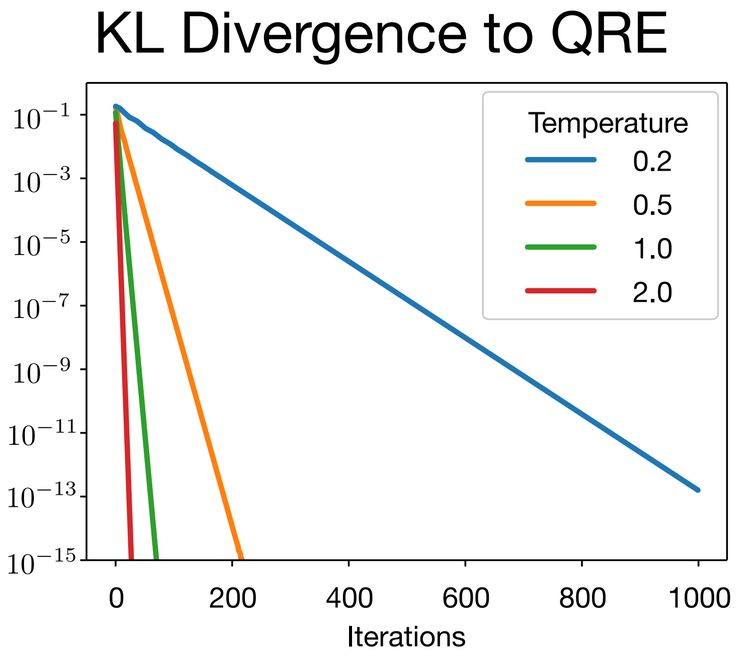
What can we say about magnetic mirror descent?

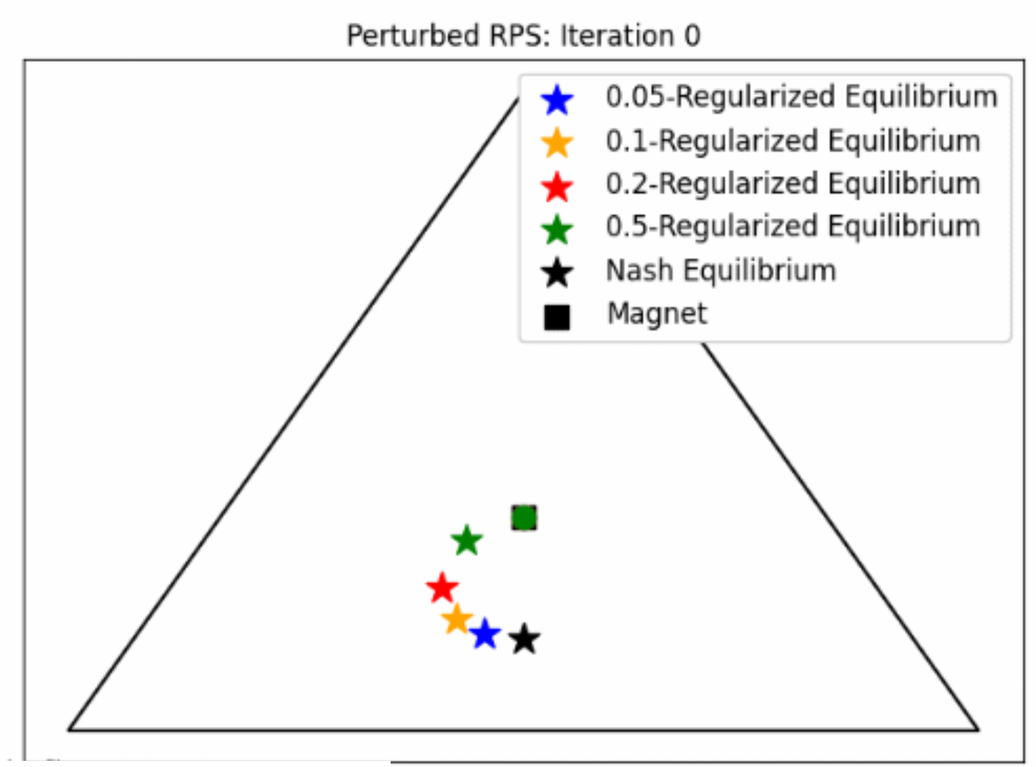
In two-player zero-sum normal-form games, if $\eta \leq \alpha / L^2$ magnetic mirror descent converges exponentially fast to a regularized equilibrium in self-play.



What can we say about magnetic mirror descent?

In two-player zero-sum normal-form games, if $\eta \leq \alpha / L^2$ magnetic mirror descent converges exponentially fast to a regularized equilibrium in self-play.



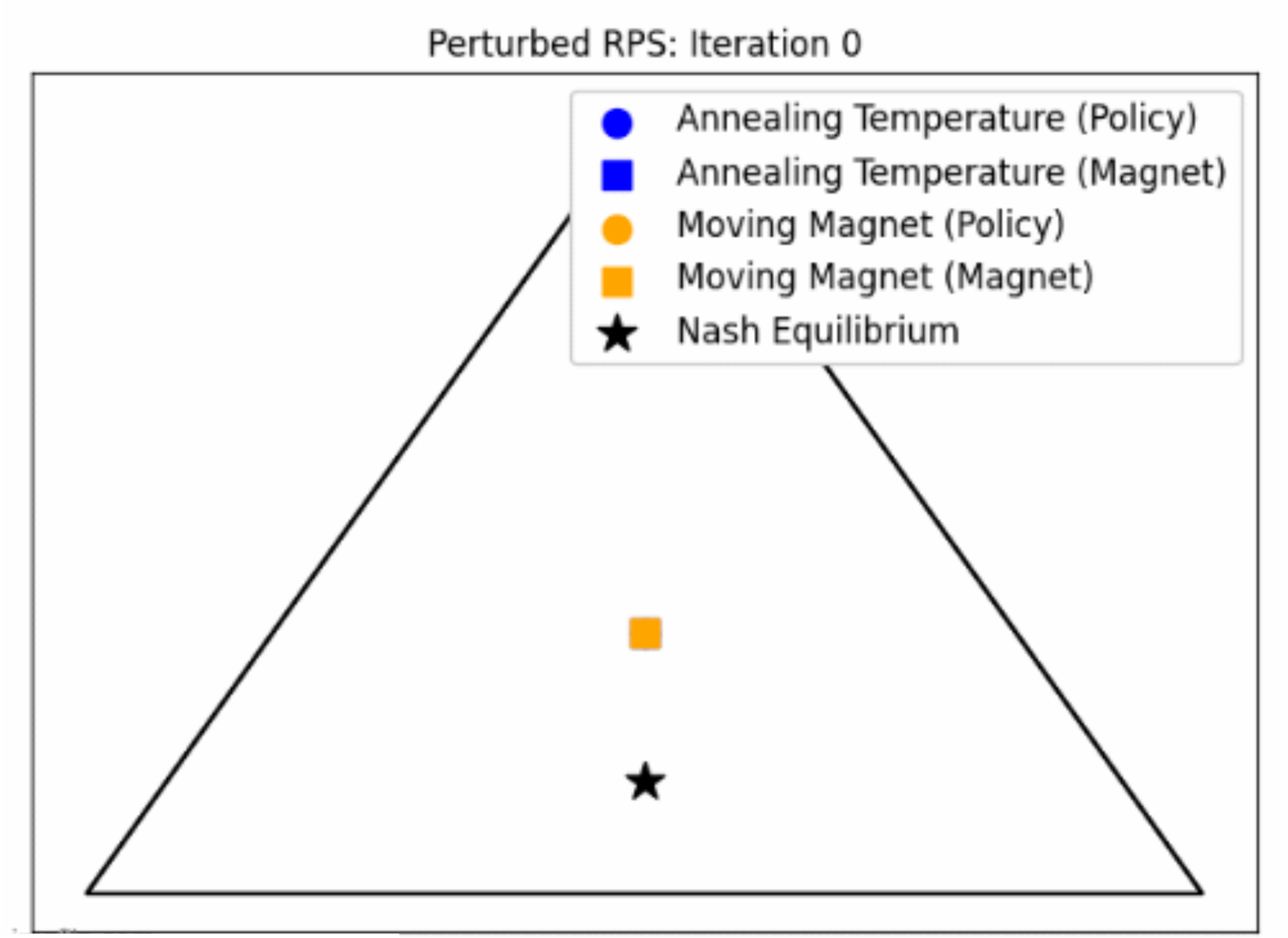


What can we say about magnetic mirror descent?

What about Nash equilibria?

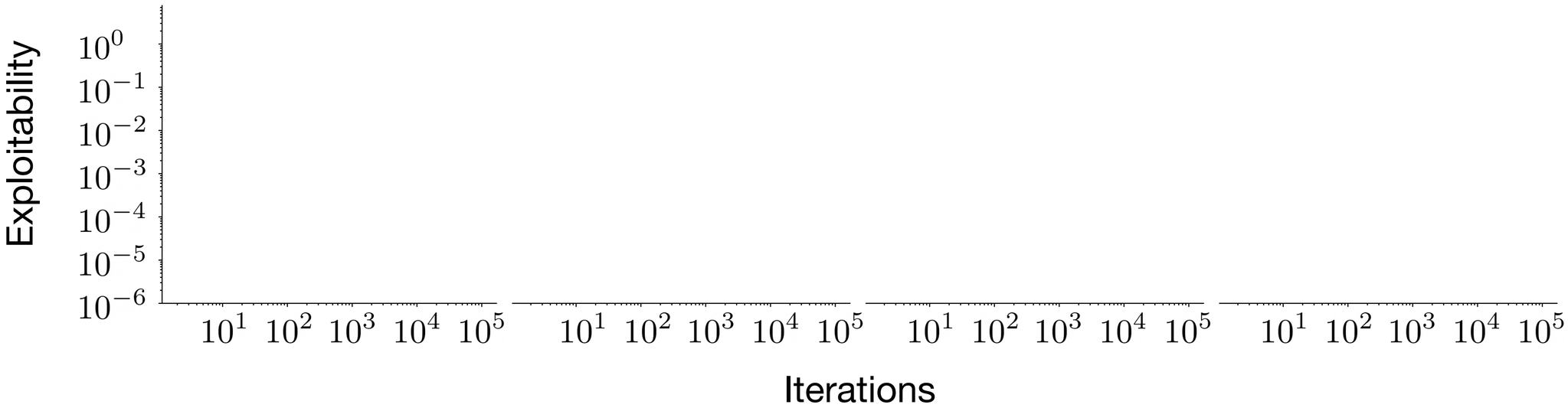
What can we say about magnetic mirror descent?

What about Nash equilibria?



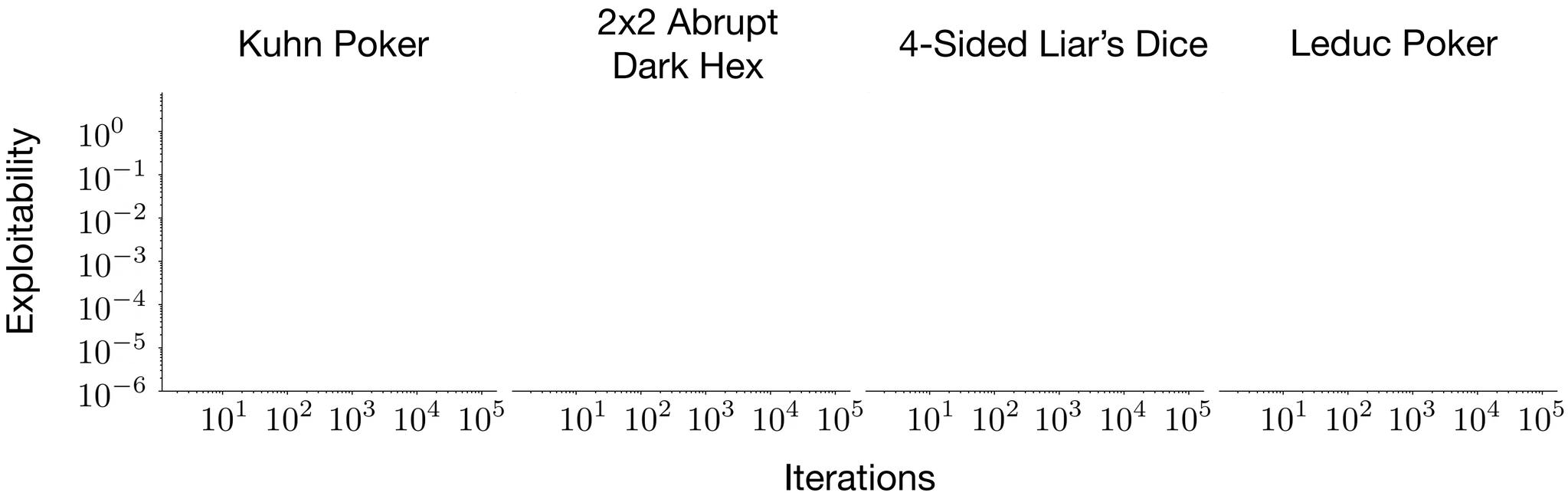
What can we say about magnetic mirror descent?

Does it converge in extensive-form games?



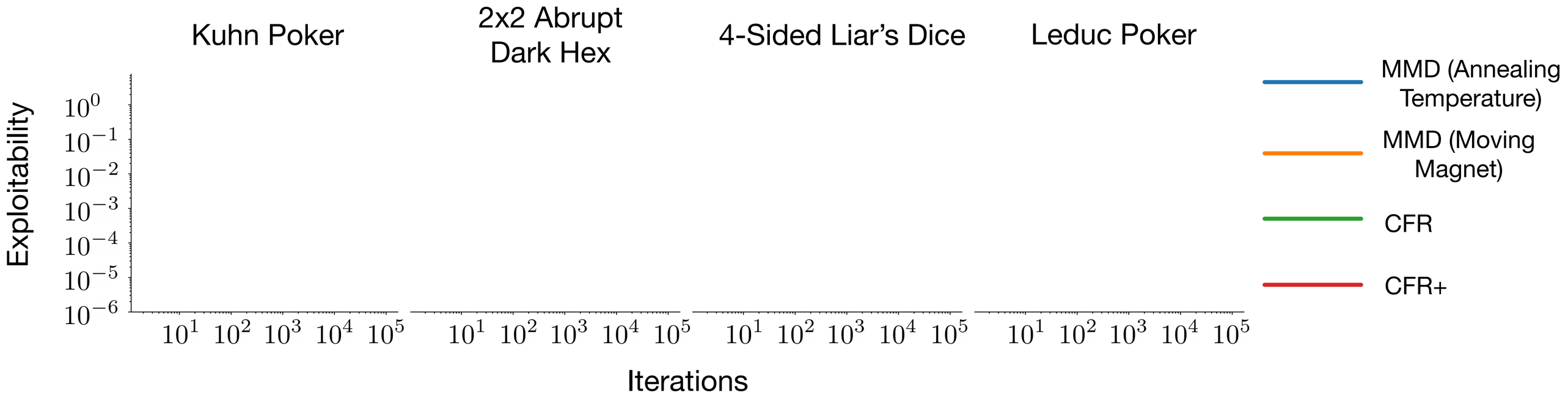
What can we say about magnetic mirror descent?

Does it converge in extensive-form games?



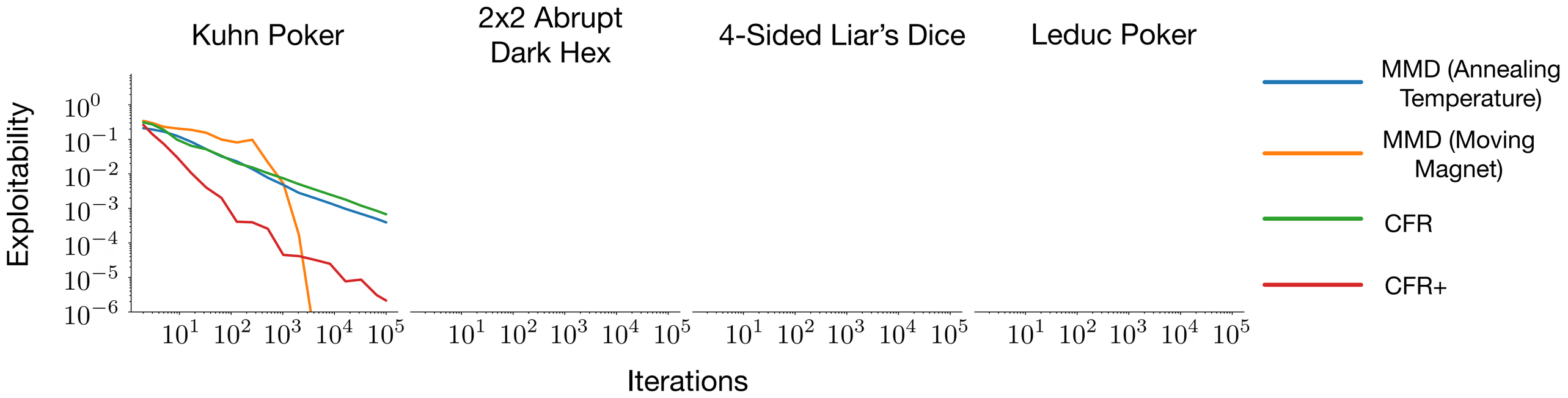
What can we say about magnetic mirror descent?

Does it converge in extensive-form games?



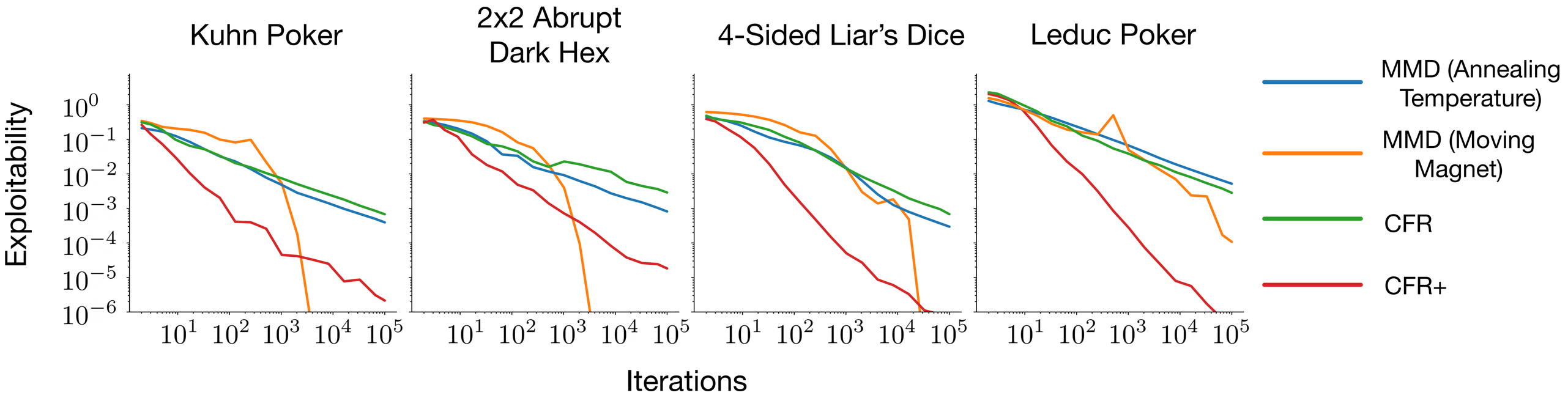
What can we say about magnetic mirror descent?

Does it converge in extensive-form games?



What can we say about magnetic mirror descent?

Does it converge in extensive-form games?



“Game-Theorified” Deep RL Approaches

Pros and Cons

“Game-Theorified” Deep RL Approaches

Pros and Cons

- Shakier theoretical foundation

“Game-Theorified” Deep RL Approaches

Pros and Cons

- Shakier theoretical foundation

+ Scale naturally

TLDR

Some approaches for large imperfect information games

TLDR

Some approaches for large imperfect information games

1. Use deep reinforcement learning to approximate best response for fictitious play or double oracle.

TLDR

Some approaches for large imperfect information games

1. Use deep reinforcement learning to approximate best response for fictitious play or double oracle.

2. Use deep learning to approximate regret values for CFR.

TLDR

Some approaches for large imperfect information games

1. Use deep reinforcement learning to approximate best response for fictitious play or double oracle.

2. Use deep learning to approximate regret values for CFR.

3. Use regularized deep policy gradient algorithms.