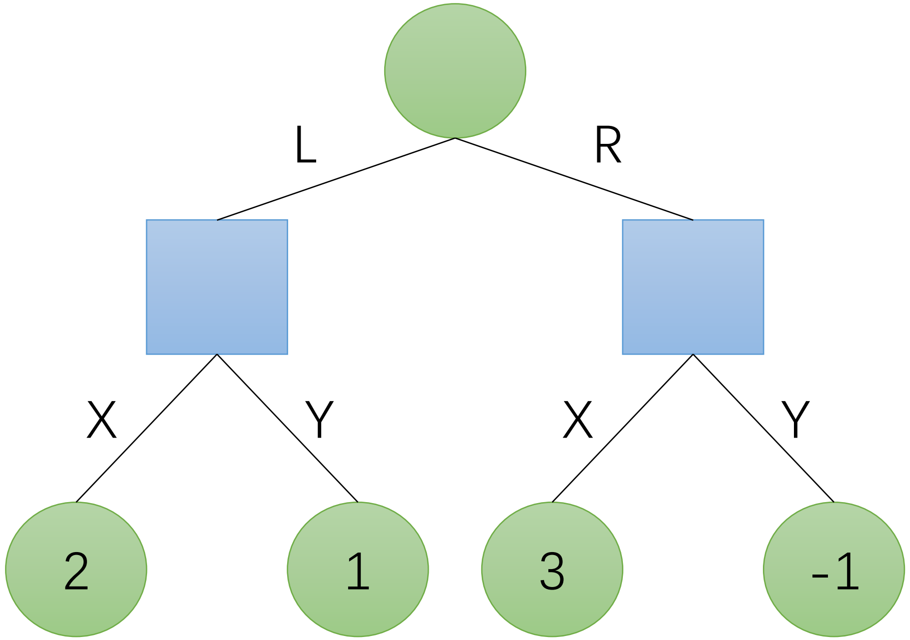
Regret-Minimizing Double Oracle for Extensive-Form Games

Xiaohang Tang,

Department of Statistical Science, UCL

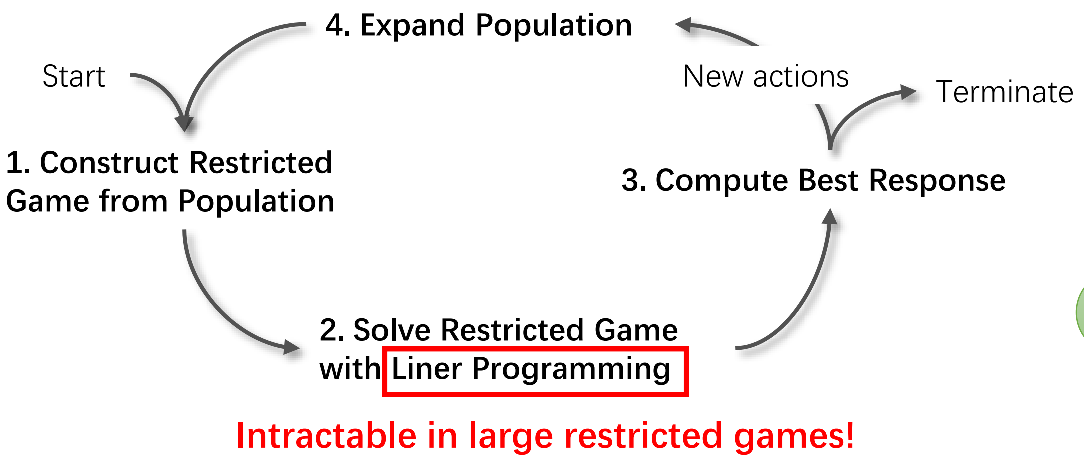
Introduction

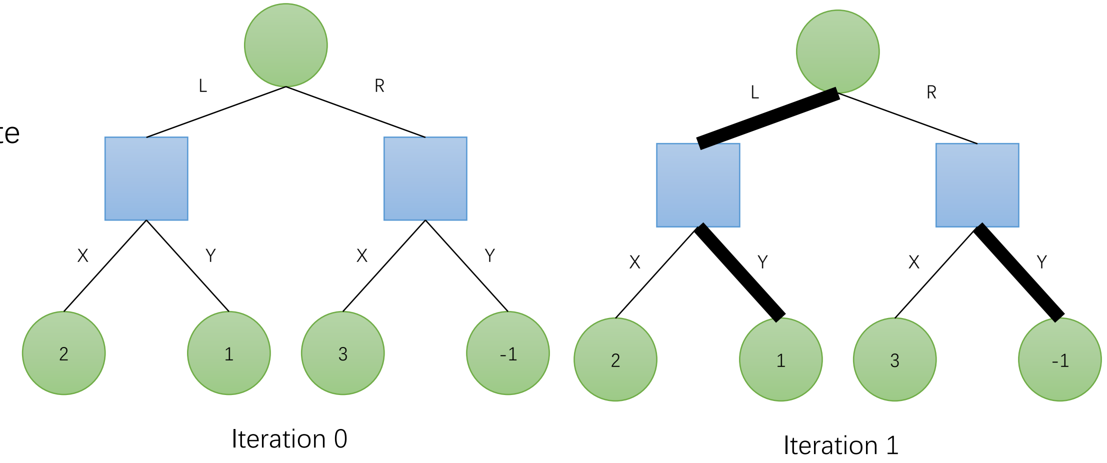
- Extensive-form Games
  - Model of multi-agent sequential decision-making problems.
  - **Solution:** (approximate) Nash Equilibrium
    - Linear Programming
    - Regret minimization
    - **Double Oracle**



Introduction

- **Double Oracle**
  - Iteratively expand restricted game with Best Response.
  - Then the NE of the restricted game is the solution.

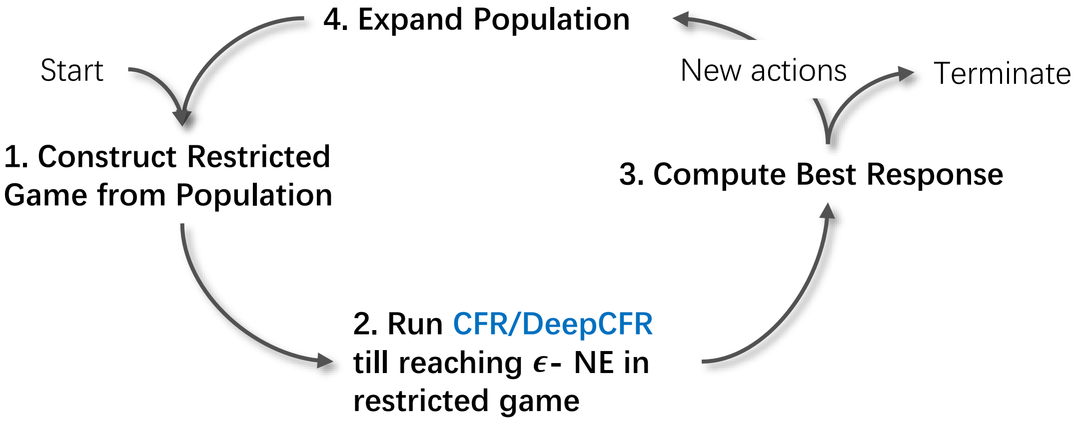




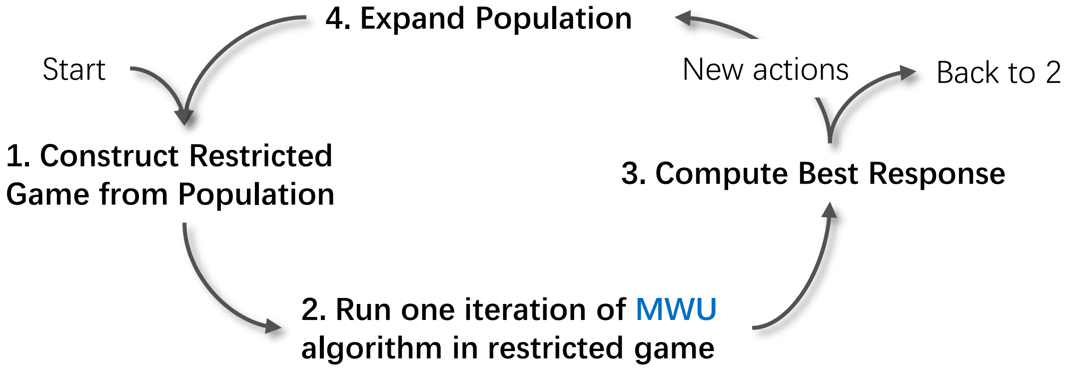
McMahan, H. Brendan, Geoffrey J. Gordon, and Avrim Blum. "Planning in the presence of cost functions controlled by an adversary." *ICML* (2003).

Introduction

- **Double Oracle + Regret Minimization**



XDO [McAleer, et al. NIPS (2021)]



ODO [Dinh, et al. TMLR (2022)]

Introduction

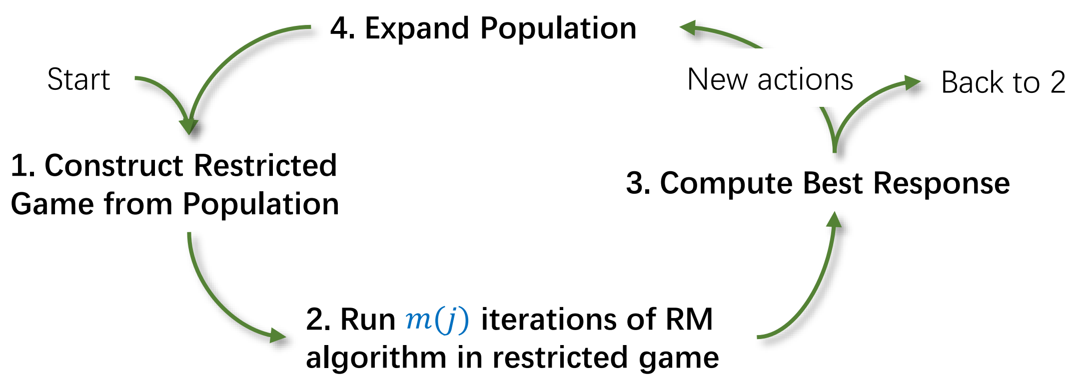
Double Oracle combined with regret minimization has shown rapid convergence to Nash Equilibrium in Games.

Main problem: a comprehensive analysis of their convergence rate or sample complexity is still lacking.

McAleer, Stephen, et al. "XDO: A double oracle algorithm for extensive-form games." *NeurIPS* (2021): 23128-23139.
Bosansky, Branislav, et al. "An exact double-oracle algorithm for zero-sum extensive-form games with imperfect information." *JAIR* (2014).

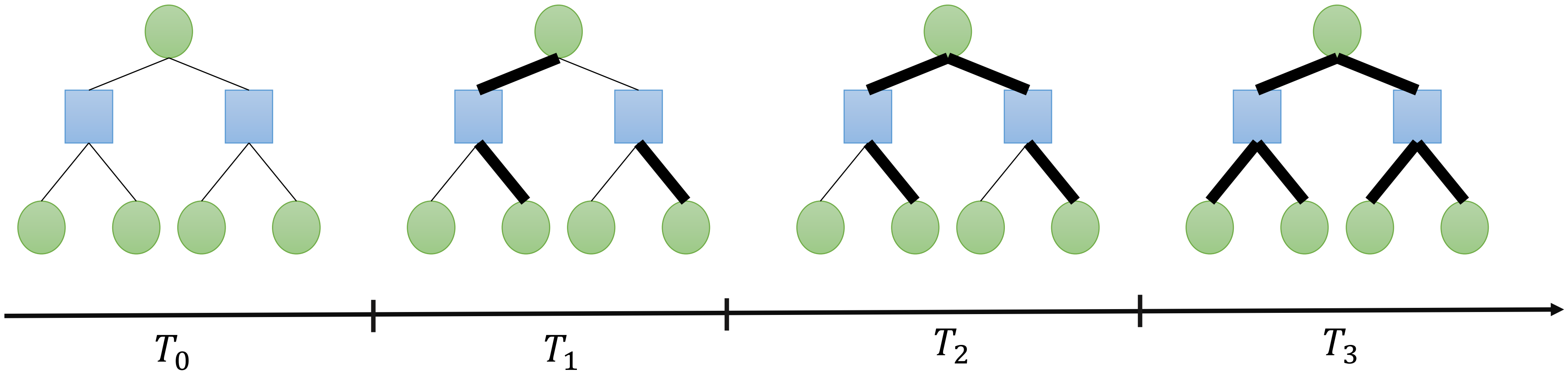
Regret-Minimizing Double Oracle

We propose a novel generic Double Oracle framework RMDO to analyze the sample complexities. Our framework employs a **frequency function $\boldsymbol{m}(\cdot)$** to represent the iterations of RM (CFR+) executed on each restricted game.



Regret-Minimizing Double Oracle

- **Frequency function** of BR computing



- Time window T_j : partition of iterations satisfying the same restricted game in the same time window. *[Dinh, et al. TMLR (2022)]*
  - **Lemma 1 (Informal***) In RMDO, suppose there are* k *windows. Denote* S *as the set of all information states, then* $\mathbf{k} \leq |\boldsymbol{S}|$**.**
- Frequency function $m(j)$:

$$N^+ \cap [0, k] \rightarrow N^+,$$

indicating that in window T_j we compute BR in every $m(j)$ iterations of RM.

Regret-Minimizing Double Oracle

Overall Average Strategy (OAS)

- **Theorem 1:** *The average regret of RMDO converges to 0 if* $m(j)$ *is sublinear:*

$$\frac{R_i(T)}{T} \leq \tilde{O}(\sum_{j=0}^{k-2} \frac{T_j}{T} \cdot [m(j) - 1] + \sum_{j=0}^{k-1} \frac{\sqrt{k}|S_i||T_j|}{T\sqrt{\{|T_j| - m(j) + 1\}}})$$

Last-window Average Strategy (LAS)

- **Theorem 2:** *Expected iterations to reach* ϵ*-NE:*

$$\tilde{O}(\frac{k|A||S|^2}{\epsilon^2} - k + \sum_j m(j))$$

Schemes of Frequency Function:

ODO: OAS of RMDO with $m(j) = 1$, sample complexity $\tilde{\boldsymbol{O}}\left(\frac{2|S|^3 k^2}{\epsilon^2}\right)$

XDO: LAS of RMDO with $m(j) \geq \frac{4^j |S_{i,j}|^2 |A_{\{i,j\}}|}{\epsilon_0^2}$, sample complexity $\tilde{\boldsymbol{O}}\left(\frac{k|A||S|^3}{\epsilon^2} + \frac{\boxed{4^k}|A||S|^3}{\epsilon_0}\right)$

Can be exponential in $|S|$! (Lemma 1)

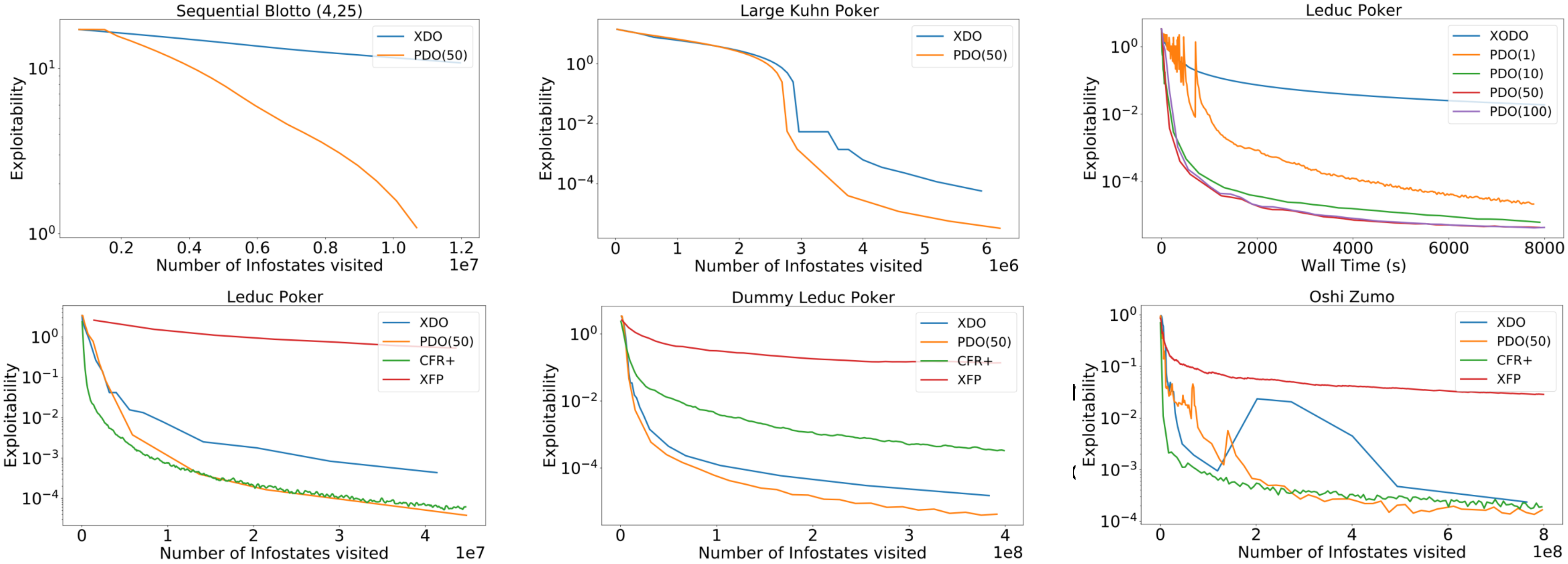
Regret-Minimizing Double Oracle

- **Periodic Double Oracle**
  - $m(j) = c\ (> 1)$, RMDO computes BR less frequent than ODO but its **last-window** average strategy achieve the least sample complexity.

EXAMPLE	SAMPLE COMPLEXITY	SAMPLE COMPLEXITY IN k
XODO	$\mathcal{O}(2\|S\|^3 k^2/\epsilon^2)$	POLYNOMIAL
XDO	$\mathcal{O}(k\|A\|\|S\|^3/\epsilon^2 + \|A\|\|S\|^3 4^k/\epsilon_0^2)$	EXPONENTIAL
PDO	$\mathcal{O}(k\|A\|\|S\|^3/\epsilon^2 + ck\|S\| + k\|A\|\|S\|^3/c\epsilon^2 - k\|S\|/c)$	LINEAR

Summary of instances of RMDO, $k \leq |S|$.

Regret-Minimizing Double Oracle



- PDO significantly outperforms XDO and XODO, highlighting its remarkable sample efficiency. Meanwhile it is competitive with SOTA such as CFR methods.
- PDO exhibits the best performance in Dummy Leduc Poker, confirming its inheritance of the advantages of DO, namely its ability to rapidly solve games with small support NE.

Regret-Minimizing Double Oracle

- We introduce a novel theoretical framework, **RMDO**, to analyze the sample complexity of the double oracle in Extensive-Form Games (EFGs).
- Based on RMDO, we extend ODO to address EFGs and reveal that the sample complexity of XDO (SOTA) can be exponential in the number of information sets.
- Then we propose a more sample-efficient algorithm **Periodic DO (PDO)** and show its fast convergence in experiments.