Regularization for Strategy Exploration in Empirical Game-Theoretic Analysis

Extended Abstract

Yongzhao Wang
University of Michigan
Ann Arbor, USA
wangyzh@umich.edu

Michael P. Wellman
University of Michigan
Ann Arbor, USA
wellman@umich.edu

ABSTRACT

We propose a novel meta-strategy solver called *regularized replicator dynamics* (RRD) for empirical game-theoretic analysis and show that RRD outperforms existing meta strategy solvers in various games.

KEYWORDS

Empirical Game-Theoretic Analysis; Policy Space Response Oracle; Regularization

ACM Reference Format:
Yongzhao Wang and Michael P. Wellman. 2023. Regularization for Strategy Exploration in Empirical Game-Theoretic Analysis: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

The methodology of ***empirical game-theoretic analysis*** (EGTA) [5, 6] provides a broad toolbox of techniques for game reasoning with models based on simulation data[1]. As many multiagent systems of interest are not easily expressed or tackled analytically, EGTA offers an alternative approach whereby a space of strategies is examined through simulation, combined with game model induction and inference. The challenge of efficiently assembling an effective game model of strategies for EGTA is called the ***strategy exploration*** problem [3].

Strategy exploration in EGTA is most clearly formulated within an iterative procedure, whereby generation of new strategies is interleaved with game model estimation and analysis. The ***Policy Space Response Oracle*** (PSRO) algorithm of Lanctot et al. [4] provides a flexible framework for iterative EGTA, where at each iteration, new strategies are generated through reinforcement learning (RL). In PSRO, the component that derives the best response target is called a ***meta-strategy solver*** (MSS), as it takes an empirical game model as input and "solves" it to produce the target profile.

In this study, we adopt an explicit regularization perspective to the specification and analysis of MSSs. We propose a novel MSS called ***regularized replicator dynamics*** (RRD), which truncates the NE search process in intermediate game models based on a regret criterion. As the size of a payoff matrix is exponential in the number of players, the cost of maintaining completely specified models over the iterations of PSRO can be prohibitive beyond two players. To mitigate this issue, we employ a PSRO-compatible profile search method, called ***backward profile search*** (BPS), which finds solution concepts without simulating the whole payoff matrix.

[1]The full version of this paper is available on arxiv: http://arxiv.org/abs/2302.04928.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

2 REGULARIZATION FOR STRATEGY EXPLORATION

2.1 Regularized Replicator Dynamics

Our new MSS, called *regularized RD* (RRD) (Algorithm 1), simply runs RD on the empirical game, stopping when the regret of the current profile (w.r.t the empirical game) meets a specified regret threshold λ, or a maximum number of iterations is reached.

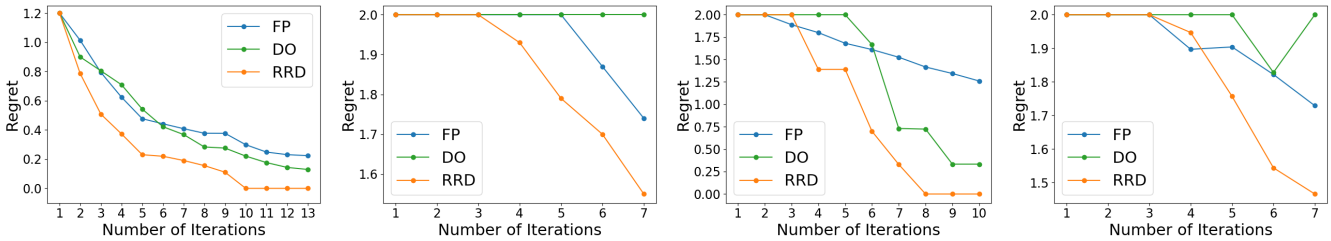
Note that RRD supports direct control of the degree of regularization through an explicit parameter: the regret threshold. This parameter is meaningful across games with different strategy sets, as long as the utility scales on which regret is measured are comparable.

Algorithm 1 RRD

Input: an empirical game $\hat{\mathcal{G}}_{S\downarrow X} = ([N], (X_i), (u_i))$
Parameters: regret threshold λ, step size α, max iterations M
Initialize RD with $\sigma_i \leftarrow \text{Uniform}(X_i)$
while regret of σ w.r.t $\hat{\mathcal{G}}_{S\downarrow X}$ $\rho^{\hat{\mathcal{G}}_{S\downarrow X}}(\sigma) > \lambda$ **do**
  for player $i \in [N]$ **do**
    $\sigma_i \leftarrow \text{Projection\_to\_simplex}(\sigma_i + \alpha \frac{d\sigma_i}{dt})$
  end for
end while
Return σ

2.2 Regularized Strategy Exploration in Multi-Player Games

We provide a simplified version of the profile search for PSRO, called *backward profile search*, and combine it with RRD to reduce the simulation cost. Distinct from the work by Brinkman and Wellman [1], BPS starts search from the singleton profile constituted by the newest-added strategies in empirical game at the current PSRO iteration. Then BPS searches potential deviations back to strategies from previous PSRO iterations. Once BPS confirms a NE of the empirical game, we applied RRD to the subgame that contains the empirical NE rather than the whole empirical game payoff matrix. In our experiments, we show that the combination of BPS and RRD can successfully find an effective best-response target in a 3-player game, without simulating the whole payoff matrix.



(a) Hex. (b) Connect four. (c) Misere Tic Tac Toe. (d) Go (size=4).

Figure 1 RRD performance compared to FP and DO in four games studied by Czarnecki et al. [2].

3 EXPERIMENTAL RESULTS

3.1 Experimental Results

3.1.1 2-player Leduc Poker. In Figure 2, we test our algorithm on 2-player Leduc poker. We first observe that RRD yields a rapid convergence to a low-regret value compared to other MSSs.

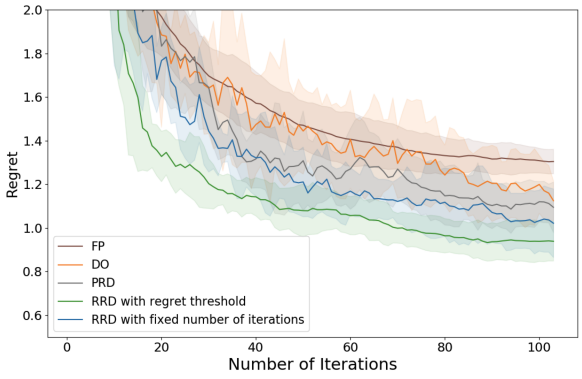


Figure 2 RRD performance in 2-player Leduc Poker.

3.1.2 Real-World Games. We further evaluate our algorithms in four of the "real-world games" studied by Czarnecki et al. [2]: Hex, Connect four, Misere Tic Tac Toe, and Go. We observe that RRD exhibits faster convergence than FP and DO in all four games.

3.1.3 Multi-Player Games. We apply the combination of BPS and RRD to 3-player Leduc poker. As shown in Figure 3, although RRD is only applied to the subgame of the empirical game, learning still benefits from regularization.

3.1.4 Attack-Graph Games. An *attack-graph game* is a two-player general-sum game defined on the attack graph where an attacker attempts to compromise a sequence of nodes to reach reach *goal* nodes and a defender endeavors to protect any node (e.g., deny an access). From Figure 4, we observe that even though the game of interest is large and beyond two-player zero-sum, RRD exhibits faster convergence and less variance than DO and FP.

3.2 A Novel Explanation for RRD Performance

Our key insight is that the performance of strategy exploration is strongly related to the regret of best-response targets *w.r.t the full*

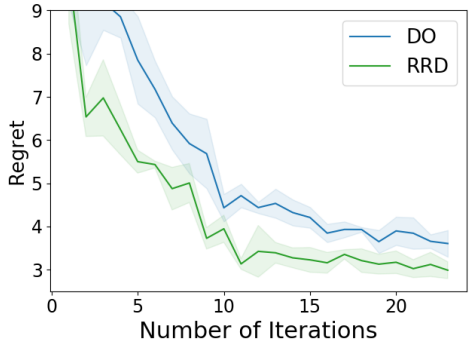


Figure 3 RRD performance in 3-player Leduc poker.

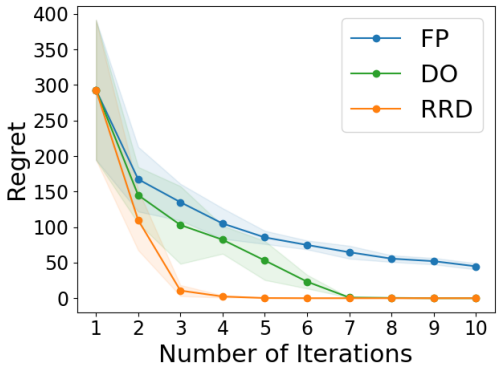


Figure 4 RRD outperforms FP and DO in the attack-graph game.

game. We note that throughout runs of PSRO, the regret of the RRD solution is much smaller than that of the empirical NE. In other words, whereas RRD has higher regret than NE in the empirical game (λ versus zero), it reliably has lower regret in the full game. Since our ultimate objective is a full-game low-regret solution, this helps to explain why the regularization imposed by RRD apparently provides robustly improved performance for strategy exploration.

REFERENCES

[1] Erik Brinkman and Michael P. Wellman. 2016. Shading and efficiency in limit-order markets. In *IJCAI-16 Workshop on Algorithmic Game Theory*.

[2] Wojciech Marian Czarnecki, Gauthier Gidel, Brendan Tracey, Karl Tuyls, Shayegan Omidshafiei, David Balduzzi, and Max Jaderberg. 2020. Real World Games Look Like Spinning Tops. In *34th Conference on Neural Information Processing Systems*.

[3] Patrick R. Jordan, L. Julian Schvartzman, and Michael P. Wellman. 2010. Strategy Exploration in Empirical Games. In *9th International Conference on Autonomous Agents and Multi-Agent Systems* (Toronto). 1131–1138.

[4] Marc Lanctot, Vinicius Zambaldi, Audrūnas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. 2017. A unified game-theoretic approach to multiagent reinforcement learning. In *31st Annual Conference on Neural Information Processing Systems* (Long Beach, CA). 4190–4203.

[5] Karl Tuyls, Julien Pérolat, Marc Lanctot, Edward Hughes, Richard Everett, Joel Z. Leibo, Csaba Szepesvári, and Thore Graepel. 2020. Bounds and dynamics for empirical game theoretic analysis. *Autonomous Agents and Multi-Agent Systems* 34 (2020), 7.

[6] Michael P. Wellman. 2016. Putting the agent in agent-based modeling. *Autonomous Agents and Multi-Agent Systems* 30 (2016), 1175–1189.