

Regret Minimization with Adaptive Opponents in Repeated Games

Mingyang Liu[‡]

LIDS, EECS, Massachusetts Institute of Technology

LIUMY19@MIT.EDU

Asuman Ozdaglar

LIDS, EECS, Massachusetts Institute of Technology

ASUMAN@MIT.EDU

Tiancheng Yu

OpenAI

YUTC14@GMAIL.COM

Kaiqing Zhang

University of Maryland, College Park

KAIQING@UMD.EDU

Editors: Steve Hanneke and Tor Lattimore

Abstract

In this paper, we study regret minimization in repeated games with *adaptive* opponents whose strategies may depend on the histories of play. The classical online learning metric of *external regret* does not fully capture such adaptivity, since it compares against decisions while treating the loss sequence as fixed. To account for the counterfactual reasoning of the players, we introduce a new metric, Repeated Policy Regret (RP-Regret), specific to this game-theoretic setting, which measures the difference between the *realized* and the *best-in-hindsight* accumulated utility when all players can *respond* to the history of play. Compared with existing regret notions in adaptive environments, RP-Regret allows stronger dynamic comparators and less restricted opponents, while still enabling the learning of better equilibria when all players minimize it.

We first identify necessary conditions for achieving sublinear RP-Regret. The comparator strategies must have sublinear accumulated variation, and both the comparator and the opponents must have imperfect recall (Piccione and Rubinstein, 1997; Waugh et al., 2009). Without these conditions, sublinear RP-Regret is impossible to achieve in general.

We then provide additional sufficient conditions and algorithms for minimizing RP-Regret. A key challenge is that RP-Regret is *nonconvex* in the strategy space by definition. We address this challenge through three approaches. The first approach uses a nonconvex optimization oracle, as in prior work on online nonconvex learning (Suggala and Netrapalli, 2020). The second approach minimizes a convex *linearized* surrogate at each iteration, which yields the minimization of a local variant of RP-Regret. The third approach directly minimizes RP-Regret when the opponents change their strategies slowly, by reformulating the repeated game as a Markov game and optimizing over occupancy measures.

Finally, we show that when all players can run algorithms to minimize the RP-Regret or its linearized variant, certain subgame-perfect equilibria of the repeated game can be learned. We also provide experiments to show that these regret notions can lead to more cooperative outcomes with higher utility in games such as the Stag Hunt.¹

Keywords: Repeated Games; Regret Minimization; Adaptive Opponents; Learning in Games

[‡] Alphabetical Order

1. Extended abstract. Full version appears as [arXiv:2606.06486v1].

References

- Michele Piccione and Ariel Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1):3–24, 1997.
- Arun Sai Suggala and Praneeth Netrapalli. Online non-convex learning: Following the perturbed leader is optimal. In *Algorithmic Learning Theory*, pages 845–861. PMLR, 2020.
- Kevin Waugh, Martin Zinkevich, Michael Johanson, Morgan Kan, David Schnizlein, and Michael H Bowling. A practical use of imperfect recall. In *SARA*, 2009.