Journal of Economic Behavior & Organization 81 (2012) 153–158

Contents lists available at SciVerse ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

On the impossibility of achieving no regrets in repeated games☆

Karl Schlag[a], Andriy Zapechelnyuk[b,*]

[a] Department of Economics, University of Vienna, Hohenstaufengasse 9, 1010 Vienna, Austria
[b] School of Economics and Finance, Queen Mary, University of London, Mile End Road, London E1 4NS, UK

ARTICLE INFO

Article history:
Received 8 March 2011
Received in revised form
21 September 2011
Accepted 22 September 2011
Available online 1 October 2011

JEL classification:
C73
D81
D83

Keywords:
Adaptive strategies
Regret minimization
Regret matching

ABSTRACT

Regret-minimizing strategies for repeated games have been receiving increasing attention in the literature. These are simple adaptive behavior rules that lead to no regrets and, if followed by all players, exhibit nice convergence properties: the average play converges to correlated equilibrium, or even to Nash equilibrium in certain classes of games. However, the no-regret property relies on a strong assumption that each player treats her opponents as unresponsive and fully ignores the opponents' possible reactions to her actions. We show that if at least one player is slightly responsive, it is impossible to achieve no regrets, and convergence results for regret minimization with responsive opponents are unknown.

1. Introduction

In a repeated interaction, an individual follows a *regret-minimizing* strategy if, loosely speaking, she reinforces those actions that she regrets not having played enough in the past. A particularly simple strategy is *regret matching*, which is defined by the following rule:

> Switch next period to a different action with a probability that is *proportional* to the *regret* for that action, where regret is defined as the increase in payoff had such a change always been made in the past (Hart and Mas-Colell, 2000; Hart, 2005).

This strategy, in particular, has the property "never change a winning team," in other words, do not switch to a different action, as long as the current action keeps being a best reply to the observed (average) actions of the opponents.

Regret-minimizing strategies that lead to "no regrets" irrespective of what the opponents play, called *no-regret strategies*, received a lot of attention in the recent literature.[1] The main value of these strategies is that they are simple adaptive behavior rules that are neither computationally demanding nor relying on common knowledge assumptions and yet exhibiting nice convergence properties. If all players follow no-regret strategies, their average joint play converges to the set of correlated

☆ We thank Jörg Oechssler and an anonymous referee for valuable comments that led to a significant improvement of the paper.

* Corresponding author.
E-mail addresses: karl.schlag@univie.ac.at (K. Schlag), a.zapechelnyuk@qmul.ac.uk (A. Zapechelnyuk).

[1] A non-exhaustive list includes Littlestone and Warmuth (1994), Fudenberg and Levine (1995), Foster and Vohra (1998, 1999), Freund and Schapire (1999), Hart and Mas-Colell (2000, 2001, 2003), Lehrer (2003), Young (2004), Cesa-Bianchi and Lugosi (2003, 2006), Lehrer and Solan (2009).

0167-2681/$ – see front matter
doi:10.1016/j.jebo.2011.09.014

equilibria or to the Hannan set,[2] depending on the notion of regret in use (Hart and Mas-Colell, 2000; see also Lehrer, 2003; Cesa-Bianchi and Lugosi, 2006); or even to Nash equilibria on certain classes of games (Hart and Mas-Colell, 2003; Marden et al., 2007).

In this note we raise the question of validity of the regret minimization objective in the context of games. On the one hand, according to the notions of regret used in the literature, an individual who contemplates whether she could have done better by having played a particular action more often in the past does not take into account the effect of her actions on the subsequent behavior of her opponent. This is perfectly fine in a decision making environment, but *not* in a game, where, by definition, players are responsive to their opponents' behavior. We show by example that failure to take the opponent's responsiveness into account may lead to unrealistic behavior.[3]

On the other hand, if we extend the notion of regret to take into account the above mentioned effect, then it becomes impossible to guarantee no regrets, even against a severely restricted set of the opponent's strategies. We show that if an opponent is slightly responsive to the player's past behavior, the maximum regret need not converge to zero. Consequently, even if all players play regret-minimizing strategies (such as Hart and Mas-Colell's (2000) regret matching) with respect to this extended notion of regret, their regrets need not vanish in the long run, and consequently, the known convergence results are not guaranteed.

2. Regrets

Consider a finite two-player game, with players named Alice and Bob.[4] Let A and B be sets of actions of Alice and Bob, respectively, and let $u : A \times B \to \mathbb{R}$ be Alice's payoff function. The game is played repeatedly in time periods $t = 1, 2, \ldots$, in which players choose actions (a_t, b_t). The history of realized actions is observable for both players.

Denote by $\overline{U}_T(a, b)$ the average payoff of Alice up to period T,

$$\overline{U}_T(a, b) = \frac{1}{T}\sum_{t=1}^{T} u(a_t, b_t),$$

and denote by $U_T(a_{(a^*|a')}, b)$ the average payoff that Alice would have obtained had she played a' instead of the reference action a^* every time in the past when she actually played a^*,

$$U_T(a_{(a^*|a')}, b) = \frac{1}{T}\sum_{t=1}^{T} w_t(a'),$$

where

$$w_t(a') = \begin{cases} u(a', b_t), & \text{if } a_t = a^*, \\ u(a_t, b_t), & \text{if } a_t \neq a^*. \end{cases}$$

Alice's *regret* $r_T(a', a^*; a, b)$ for choosing action a^* instead of action a' after T periods is defined as the excess of $U_T(a_{(a^*|a')}, b)$ over $\overline{U}_T(a, b)$,

$$r_T(a', a^*; a, b) = U_T(a_{(a^*|a')}, b) - \overline{U}_T(a, b).$$

The objective of the previous literature has been to identify strategies for Alice that guarantee no regrets in the long run. More specifically, let $h_T = ((a_1, b_1), \ldots, (a_T, b_T))$ denote the history of play up to T, and let $\mathcal{H}$ be the set of all finite histories. Alice has a no-regret strategy $\alpha : \mathcal{H} \to \Delta(A)$ if $\limsup_{T\to\infty} r_T(a', a^*; a, b) \leq 0$ holds almost surely under α for all deterministic sequences b and all pairs of actions (a^*, a').

According to the above definition of regret,[5] Alice evaluates her regret for choosing action a^* instead of action a' by contemplating how much higher payoff, on average, she could have received had she played a' in every past period when she actually played a^*, assuming that *the play of the opponents would have remained unchanged*. This definition is plausible in the context of decision making, when an individual's actions have no effect on the opponent, who can be perceived as an

[2] The Hannan set of a game is the set of all mixed action profiles that satisfy Hannan's (1957) no-regret condition. It is also known as the set of *coarse correlated equilibria* first appeared in Moulin and Vial (1978), but explicitly defined as a solution concept by Young (2004, ch. 3).

[3] This problem is recognized in the computer science literature. Farias and Megiddo (2004) and Cesa-Bianchi and Lugosi (2006, ch. 7.11) show that regret minimizing strategies fail to lead to the cooperative outcome in a repeated prisoner's dilemma. Our example is different and, as we believe, has a value on its own, as it illuminates failure to learn the Pareto dominant equilibrium of a *one-shot game*, whereas the above literature shows failure to learn playing strictly dominated actions.

[4] Bob can be considered as a set of players, so the arguments presented below trivially extend to n-player games.

[5] Specifically, we have been considering *conditional regrets*. The *unconditional regret* of Alice for an action a' refers to the difference in her average payoff had she always chosen a' instead of her actual past play. "No conditional regret" implies "no unconditional regret", but not vice versa, unless Alice has only two actions.