JMLR: Workshop and Conference Proceedings vol 49:1–2, 2016

Online Learning and Blackwell Approachability in Quitting Games

János Flesch J.FLESCH@MAASTRICHTUNIVERSITY.NL
Department of Quantitative Economics, Maastricht University, The Netherlands

Rida Laraki RIDA.LARAKI@LAMSADE.DAUPHINE.FR
C.N.R.S. at LAMSADE (University Paris Dauphine) and Economics Department (Ecole Polytechnique), Paris, France

Vianney Perchet VIANNEY.PERCHET@NORMALESUP.ORG
Laboratoire de Statistique, ENSAE - CREST, Malakoff, France

Abstract

We consider the sequential decision problem known as regret minimization, or more precisely its generalization to the vectorial or multi-criteria setup called Blackwell approachability. We assume that Nature, the decision maker, or both, might have some quitting (or terminating) actions so that the stream of payoffs is constant whenever they are chosen. We call those environments "quitting games".

We characterize convex target sets $\mathcal{C}$ that are Blackwell approachable, in the sense that the decision maker has a policy ensuring that the expected average vector payoff converges to $\mathcal{C}$ at some given horizon known in advance. Moreover, we also compare these results to the cases where the horizon is not known and show that, unlike in standard online learning literature, the necessary or sufficient conditions for the anytime version of this problem are drastically different than those for the fixed horizon.

Keywords: Online Learning, Blackwell Approachability, Markov Decision Process, Absorbing Games

Quitting games We consider the setting where both the decision maker and Nature have a finite set of actions, denoted respectively by $\mathbf{I} = \mathcal{I} \cup \mathcal{I}^\star$ and $\mathbf{J} = \mathcal{J} \cup \mathcal{J}^\star$. The actions in $\mathcal{I}$ and $\mathcal{J}$ are called non-quitting, and the actions in $\mathcal{I}^\star$ and $\mathcal{J}^\star$ are called quitting. To each pair of actions $(i,j) \in \mathbf{I} \times \mathbf{J}$ is associated a payoff vector $g(i,j) \in \mathbb{R}^d$.

The game between the decision maker and Nature is played sequentially and at stage $t \in \mathbb{N}$, they choose the actions i_t and j_t simultaneously. If only non-quitting actions have been played before stage t, i.e. $i_{t'} \in \mathcal{I}$ and $j_{t'} \in \mathcal{J}$ for every $t' < t$, then the decision maker is free to choose any action in $\mathbf{I}$ and Nature is free to choose any action in $\mathbf{J}$. However, if a quitting action was played by either player at a stage prior to stage t, i.e. $i_{t'} \in \mathcal{I}^\star$ or $j_{t'} \in \mathcal{J}^\star$ for some $t' < t$, then $i_t = i_{t-1}$ and $j_t = j_{t-1}$.

We denote by $\Delta(E)$ the set of probability measures over a set E and by $\mathcal{M}(E)$ the set of positive measures. As it is usual, the payoff mapping g is extended multi-linearly to $\Delta(\mathbf{I})$ and $\Delta(\mathbf{J})$ by $g(\mathbf{x},\mathbf{y}) = \sum_{i,j} \mathbf{x}_i \mathbf{y}_j g(i,j)$ and, more generally, to the set of measures $\mathcal{M}(\mathbf{I})$ and $\mathcal{M}(\mathbf{J})$. We also introduce the probability of absorption and the expected absorption payoff, defined respectively by $p^\star(\alpha,\beta) = \sum_{(i,j)\in(\mathcal{I}^\star\times\mathbf{J})\cup(\mathbf{I}\times\mathcal{J}^\star)} \alpha_i\beta_j$ and $g^\star(\alpha,\beta) = \sum_{(i,j)\in(\mathcal{I}^\star\times\mathbf{J})\cup(\mathbf{I}\times\mathcal{J}^\star)} \alpha_i\beta_j g(i,j)$

Given a closed and convex set $\mathcal{C} \subset \mathbb{R}^d$ and a final horizon T, the objective of the decision maker is to ensure with a strategy σ that the expected average payoff $\mathbb{E}_{\sigma,\tau}\frac{1}{T}\sum_{t=1}^T g(i_t,j_t)$ is as close to $\mathcal{C}$ as possible. The target set $\mathcal{C}$ is called approachable if the distance at stage T converges to 0 as T increases. In this general case, our main results are

SUFFICIENCY: If the following Condition (1) is satisfied, then $\mathcal{C}$ is approachable by the decision maker.

$$\max_{\mathbf{y}\in\Delta(\mathbf{J})} \min_{\mathbf{x}\in\Delta(\mathbf{I})} \inf_{\alpha\in\mathcal{M}(\mathbf{I})} \sup_{\beta\in\mathcal{M}(\mathbf{J})} d_{\mathcal{C}}\Big(\frac{g(\mathbf{x},\mathbf{y}) + g^\star(\alpha,\mathbf{y}) + g^\star(\mathbf{x},\beta)}{1 + p^\star(\alpha,\mathbf{y}) + p^\star(\mathbf{x},\beta)}\Big) = 0 \tag{1}$$

NECESSITY: If $\mathcal{C}$ is approachable by the decision maker, then the following Condition (2) is satisfied,

$$\max_{\mathbf{y}\in\Delta(\mathbf{J})}\sup_{\beta\in\mathcal{M}(\mathbf{J})}\min_{\mathbf{x}\in\Delta(\mathbf{I})}\inf_{\alpha\in\mathcal{M}(\mathbf{I})} d_{\mathcal{C}}\Big(\frac{g(\mathbf{x},\mathbf{y})+g^\star(\alpha,\mathbf{y})+g^\star(\mathbf{x},\beta)}{1+p^\star(\alpha,\mathbf{y})+p^\star(\mathbf{x},\beta)}\Big)=0 \tag{2}$$

Only one player can quit. We consider two subclasses of quitting games in which only one of the players can quit, that is: (a) either $\mathcal{J}^\star=\emptyset$ (i.e. only the decision maker can quit) or (b) $\mathcal{I}^\star=\emptyset$ (i.e. only Nature can quit). In this restrictive setting, our main result is that if $\mathcal{J}^\star=\emptyset$, Conditions (1) and (2) coincide with the Blackwell condition, and we obtain a necessary and sufficient condition for approachability:

$$\mathcal{C} \text{ is approachable} \iff \forall \mathbf{y}\in\Delta(\mathbf{J}), \exists \mathbf{x}\in\Delta(\mathbf{I}), g(\mathbf{x},\mathbf{y})\in\mathcal{C}.$$

Consequently, in this class, a closed and convex set is either approachable or excludable.

However, the equivalence between Conditions (1) and (2) is not true if it is Nature that can quit, i.e., $\mathcal{I}^\star=\emptyset$.

Anytime vs Fixed Horizon A set $\mathcal{C}\subset\mathbb{R}^d$ is called anytime-approachable if the decision maker has a strategy independent of the horizon T such that the distance of the expected average payoff to $\mathcal{C}$ decreases to 0. We prove that if only nature can quit, i.e. $\mathcal{I}^\star=\emptyset$, Condition (1) is necessary and sufficient for anytime approachability. But if only the decision maker can quit, i.e. $\mathcal{J}^\star=\emptyset$, then there are convex sets which are approachable but not anytime approachable (i.e. Condition (1) is not sufficient for anytime approachability). This illustrates that, unlike standard approachability and regret minimization, the anytime version of a criterion might be very different than its fixed horizon version in quitting games. Stated otherwise, the doubling trick does not work.

Acknowledments
Vianney Perchet received support from the ANR Project GAGA: ANR-13-JS01-0004-01, from PGMO Project Nougat and from the CNRS Project Parasol. Rida Laraki received supported from ANR-11-IDEX-0003-02 and ANR-11- LABEX-0047.