Available online at www.sciencedirect.com

ScienceDirect

Games and Economic Behavior 66 (2009) 315–325

GAMES and Economic Behavior

www.elsevier.com/locate/geb

Approachability in repeated games: Computational aspects and a Stackelberg variant ☆

Shie Mannor [a], John N. Tsitsiklis [b,*]

[a] Department of Electrical and Computer Engineering, McGill University, Montreal, PQ, H3A 2A7 Canada
[b] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 3 March 2006

Available online 11 April 2008

Abstract

We consider a finite two-player zero-sum game with vector-valued rewards. We study the question of whether a given polyhedral set D is "approachable," that is, whether Player 1 (the "decision maker") can guarantee that the long-term average reward belongs to D, for any strategy of Player 2 (the "adversary"). We examine Blackwell's necessary and sufficient conditions for approachability, and show that the problem of checking these conditions is NP-hard, even in the special case where D is a singleton. We then consider a Stackelberg variant whereby, at each stage, the adversary gets to act after observing the decision maker's action. We provide necessary and sufficient conditions for approachability, and again establish that checking these conditions is NP-hard, even when D is a singleton. On the other hand, if the dimension of the reward vector is fixed, an approximate version of these conditions can be checked in polynomial time.

JEL classification: C73; C79

1. Introduction

We consider a decision maker (Player 1, P_1) who interacts repeatedly with the environment, modeled as an adversary (Player 2, P_2). At each stage (time step), each player chooses an action from given finite sets and a vector-valued reward is realized, as a function of the pair of actions chosen. We are given a polyhedral set D, and we are interested in the question of whether there exists a strategy for P_1 under which the long-term average of the reward vector is guaranteed to belong to D, for every strategy of P_2 (in which case, we say that D is "approachable"). This problem was introduced and studied by Blackwell (1956), using the tools of what became known as "approachability theory." In particular, Blackwell established necessary and sufficient conditions for the case of a convex set D, as well as a prescription for the strategy of P_1. However, despite the pervasiveness and the renewed interest in approachability theory

☆ This research was supported by the National Science Foundation under contract ECS-0312921, and by the Natural Sciences and Engineering Research Council of Canada.

* Corresponding author.
E-mail addresses: shie.mannor@mcgill.ca (S. Mannor), jnt@mit.edu (J.N. Tsitsiklis).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2008.03.008

in the context of learning in games, the computational aspects of this theory do not seem to have been considered before. The objective of this paper is to close this gap.

We start in Section 2 with a description of the model, and background results from approachability theory (Blackwell's conditions). Even though in some cases it is fairly straightforward to check whether a set is approachable or not (see Hart and Mas-Colell, 2001 for a representative example), we show that checking Blackwell's conditions is an NP-hard problem, even in the special case where the set D is a singleton. On the other hand, when the dimension of the reward vector is fixed, we establish that the question of approachability can be decided "approximately" in time which is polynomial (though exponential in the dimension of the reward vector).

Blackwell's formulation and conditions refer to the case where, at each stage, the two players act simultaneously, without knowledge of the other player's action. In Section 3, we introduce a Stackelberg variant in which, at each stage, P_1 acts first and P_2 (the adversary) is informed of P_1's action before choosing her own action. We establish that the question of approachability is NP-hard, even in the special case where the set D is a singleton, but can be decided in polynomial time if the dimension of the reward vector is held fixed.

2. Approachability for the case of simultaneous actions

2.1. Model and background

We consider a repeated game where a decision maker wishes to guarantee that the long-term average of a vector-valued reward belongs to a prespecified target set. The stage game is a finite game involving two players, P_1 (the decision maker) and P_2 (the adversary). This naturally abstracts the case where there are multiple players and we are only concerned with the reward obtained by P_1.

The game is defined by a triple $(\mathcal{A}, \mathcal{B}, M)$ where:

(a) $\mathcal{A}$ is the finite set of actions for P_1; we will assume that $\mathcal{A} = \{1, 2, \dots, m\}$.
(b) $\mathcal{B}$ is the finite set of actions for P_2; we will assume that $\mathcal{B} = \{1, 2, \dots, n\}$.
(c) M is an $n \times m$ matrix with vector-valued entries, with $M(a, b)$ denoting the reward obtained by P_1, when P_1 chooses action $a \in \mathcal{A}$, and P_2 chooses action $b \in \mathcal{B}$; we will assume that $M(a, b) \in \mathbb{R}^k$.[1]

The game is played in stages. At each stage t, P_1 chooses an action $a_t \in A$, P_2 chooses an action $b_t \in B$, and P_1 obtains a reward $m_t = M(a_t, b_t)$. We define P_1's average reward, at time t, as

$$\hat{m}_t = \frac{1}{t} \sum_{\tau=1}^{t} m_\tau.$$

We further assume that P_1 has a prespecified target set $D \in \mathbb{R}^k$, assumed to be a polyhedron.[2] The goal of P_1 is to have the average reward $\hat{m}_t$ approach this set D, as t increases, in a sense to be made precise below.

For a finite set $\mathcal{C}$, we let $\Delta(\mathcal{C})$ be the set of all probability measures on a set $\mathcal{C}$, which is identified with the with the set of $|\mathcal{C}|$-dimensional nonnegative vectors whose entries sum to one, and which will be referred to as the set of possible mixed actions on $\mathcal{C}$. A strategy for P_1 (respectively, P_2) is a mapping from all possible histories of the form $(a_1, b_1, \dots, a_{t-1}, b_{t-1})$ to the set of mixed actions on $\mathcal{A}$ (respectively, $\mathcal{B}$). Given the strategies of the two players, we assume that the randomizations involved are all independent. We use $\|\cdot\|$ to denote the Euclidean norm in $\mathbb{R}^k$, and define the point-to-set distance $\rho(x, D) = \inf_{y \in D} \|x - y\|$. We now define formally the goal of P_1.

Definition 2.1. A set D is *approachable* if there exists a strategy σ of P_1 such that for every $\varepsilon, \delta > 0$, there exists t_0 such that for every strategy τ of P_2,

$$\mathbf{P}\Big(\sup_{t \geqslant t_0} \rho(\hat{m}_t, D) > \varepsilon\Big) < \delta. \tag{1}$$

[1] All of the subsequent development also applies to the case where the rewards are random variables, sampled independently at each time, with mean $M(a_t, b_t)$ and finite second moment. We restrict to the deterministic case for simplicity.

[2] We restrict to polyhedral sets, as opposed to the general convex sets considered by Blackwell, because we wish to focus on algorithmic aspects.