Automatica 48 (2012) 1898–1903

Contents lists available at SciVerse ScienceDirect

Automatica

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

Brief paper

Saddle points of discrete Markov zero-sum game with stopping☆

Xun Li [a], Jie Shen [b], Qingshuo Song [b,1]

[a] Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong

[b] Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 14 April 2011
Received in revised form
4 December 2011
Accepted 13 March 2012
Available online 28 June 2012

Keywords:
Saddle points
Dynamic game
Markov chain approximation
Dynamic programming principle

ABSTRACT

We study the sufficient conditions for the existence of a saddle point of a time-dependent discrete Markov zero-sum game up to a given stopping time. The stopping time is allowed to take either a finite or an infinite non-negative random variable with its associated objective function being well-defined. The result enables us to show the existence of the saddle points of discrete games constructed by Markov chain approximation of a class of stochastic differential games.

1. Introduction

Two-player stochastic differential game problems have wide applications in, for example, pursuit-evasion games, queueing systems in heavy traffic, risk-sensitive control, and constrained optimization problems, see Elliott and Kalton (1972), Fleming and McEneaney (1995), Fleming and Soner (2006), Kushner (2002), Song (2008), Song and Yin (2006) and Song, Yin, and Zhang (2008) and the references therein. The underlying process of such a game is given by a stochastic differential equation of (16) controlled by two players. Given a cost function of (17), the objective of Player 1 is to minimize, while Player 2 is to maximize the cost function. As a consequence, the value function may have two different kinds: the upper value V^+ is obtained if the Player 1 moves first (Player 2 goes last) at real time; and the lower value V^- is obtained if two players move in the reverse order. In general, the two values are different, and it is said to be a saddle point if they are equal to each other.

The sufficient conditions for the existence of the saddle points has been studied widely in the literature within different frameworks. On a fixed finite time horizon of (19), the sufficient conditions were obtained in Fleming and Souganidis (1989) and Fleming and Soner (2006) by comparing two Hamilton–Bellman–Jacobi (HJB) equations associated with the upper and lower values. More recently, Kushner (2002) provided purely probabilistic analysis on the sufficient condition of the existence of a saddle point on the stochastic differential game with infinite time horizon of (18). The main ingredient of this work is to discretize the stochastic differential game into a discrete Markov game using a Markov chain approximation method with a step size h, and show its limit behavior $\lim_{h\to 0} V^{h,+} = \lim_{h\to 0} V^{h,-}$ of the upper value $V^{h,+}$ and lower value $V^{h,-}$ for the discrete game. Along the same line, Song and Yin (2006) and Song et al. (2008) generalized the probabilistic approach of Kushner (2002) to a regime-switching model with exit time of (20), and showed a stronger result: there exists a saddle point for each discrete game $V^{h,+} = V^{h,-}$ for all small enough h.

In this work, we will study the sufficient condition for the existence of a saddle point of a discrete Markov game in a more general setup. Given a controlled Markov chain, two players want to minimize/maximize the cost function of (1) up to a stopping time T. Our result shows that, extending a result on static game by the dynamic programming principle (DPP), there exists a saddle point when the system satisfies either convex-concavity or separability conditions, see Theorem 5.

Note that the terminal time T of the discrete game is allowed to take any stopping time, which includes fixed time, exit time, or infinity. By virtue of our result in the discrete Markov game

☆ The research of Xun Li is supported in part by the Research Grants Council of Hong Kong No. PolyU 524109 and the research of Qingshuo Song is supported in part by the Research Grants Council of Hong Kong No. CityU 103310. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor George Yin under the direction of Editor Ian R. Petersen. We sincerely thank the Editor, the Associate Editor, and the two anonymous reviewers for their helpful comments and advice.

E-mail addresses: malixun@inet.polyu.edu.hk (X. Li), jieshen2@student.cityu.edu.hk (J. Shen), song.qingshuo@cityu.edu.hk (Q. Song).

[1] Tel.: +852 3442 2926; fax: +852 3442 0250.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.06.012

in this general setup, the discrete game from the Markov chain approximation on the aforementioned stochastic differential game in Fleming and Souganidis (1989), Fleming and Soner (2006), Kushner (2002), Song (2008), Song and Yin (2006) and Song et al. (2008) always has a saddle point, provided that the original system satisfies convex-concavity or separability. Moreover, due to the different approach based upon the monotonicity Lemma 4 of difference between upper and lower values, this result also relaxes the sufficient condition obtained by Song and Yin (2006) and Song et al. (2008), see Remark 6.

The rest of this paper is organized as follows. In Section 2, we formulate the time-dependent discrete Markov zero-sum game. Section 3 provides some related properties of the value functions and the main result on sufficient conditions for the existence of a saddle point. Section 4 extends the results to the discrete games constructed from Markov chain approximation on a stochastic differential game. Throughout this paper, K stands for a general constant and may vary in different places.

2. Formulation

Consider a two-player discrete Markov zero-sum game on a discrete-time horizon $\mathbb{T}_t = \{t, t+1, t+2, \ldots\}$ for some fixed non-negative integer t. Let $S \subset \mathbb{Z}$ be a discrete (not necessarily finite) state space of a Markov chain. Control spaces U_1 and U_2 for Player 1 and Player 2 are two compact subsets of $\mathbb{R}$. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_n\}_{n\in\mathbb{T}_t})$. Let $\{\xi_n,\ n \in \mathbb{T}_t\}$ be a controlled discrete-time Markov chain, whose time-dependent transition probabilities are controlled by a pair of sequences $\{(u_{1,n}, u_{2,n}) \in U_1 \times U_2 :\ n \in \mathbb{T}_t\}$, where $u_{i,n} \in U_i$ stands for the decision at time n by Player i.

A control policy $\{(u_{1,n}, u_{2,n}) \in U_1 \times U_2 : n \in \mathbb{T}_t\}$ for the chain $\{\xi_n : n \in \mathbb{T}_t\}$ is said to be *Markovian* if

$$\mathbb{P}\{\xi_{n+1} = y|\xi_k, u_{1,k}, u_{2,k},\ k \le n\} \stackrel{\Delta}{=} p_n(\xi_n, y|u_{1,n}, u_{2,n})$$

for a sequence of non-negative measurable functions $p_n(\cdot, \cdot|\cdot, \cdot) \in [0, 1]$ with $n \in \mathbb{T}_t$, and the set of all Markovian controls is denoted by $\mathcal{U}_t$.

Given a sequence of time-dependent non-negative running cost functions $c_n(\cdot, \cdot, \cdot) : S \times U_1 \times U_2 \mapsto \mathbb{R}$, and the terminal cost functions $g_n(\cdot) : S \mapsto \mathbb{R}$, the total cost for an initial $\xi_t = x \in S$ and a Markovian control policy $(u_1, u_2) = \{(u_{1,n}, u_{2,n}) : n \in \mathbb{T}_t\}$ is defined by

$$W(t, x, u_1, u_2) = \mathbb{E}_{t,x}^{u_1,u_2}\left[\sum_{n=t}^{T-1} e^{-\lambda_n} c_n(\xi_n, u_{1,n}, u_{2,n}) + e^{-\lambda_T} g_T(\xi_T)\right], \tag{1}$$

where $T = T^{t,x,u_1,u_2}$ is a given (t, x, u_1, u_2)-dependent stopping time w.r.t $\{\mathcal{F}_n\}$ taking values in $\mathbb{T}_t$, the discount factor $\{\lambda_n\}$ satisfies

$$0 \le \lambda_n \le \lambda_{n+1}, \quad \forall n \in \mathbb{T}_t, \tag{2}$$

and $\mathbb{E}_{t,x}^{u_1,u_2}$ is the conditional expectation given that initially $\xi_t = x$ and control (u_1, u_2). The discount factor satisfying (2) includes the following commonly used case: $\lambda_n = \lambda n$ for some $\lambda > 0$ and all $n \in \mathbb{T}_0$. Note that, if $T^{t,x,u_1,u_2} = t$, then (1) is equivalent to

$$W(t, x, u_1, u_2) = e^{-\lambda_t} g_t(x). \tag{3}$$

In the rest of paper, we impose the following conditions on functions c_n and g,

$$|c_n(x, r_1, r_2)| + |g_n(x)| \le K; \qquad g_\infty(x) = 0, \tag{4}$$

for all $(n, x, r_1, r_2) \in \mathbb{T}_0 \times S \times U_1 \times U_2$. Due to (4), if the stopping time $T = \infty$ almost surely, then the discounted terminal cost $e^{-\lambda_T} g_T(\xi_T) = 0$, and the cost function W of (1) is consistent to

$$W(t, x, u_1, u_2) = \mathbb{E}_{t,x}^{u_1,u_2}\left[\sum_{n=t}^{\infty} e^{-\lambda_n} c_n(\xi_n, u_{1,n}, u_{2,n})\right]. \tag{5}$$

In this discrete Markov game, Player 1 wants to minimize, while Player 2 wants to maximize the cost. The two players have different information available depending on who makes the decision first (or who "goes first"). Using $\mathcal{U}_{t,i}(1)$, we denote the space of the admissible strategies that player i goes first starting from given time t. Similarly, $\mathcal{U}_{t,i}(2)$ stands for the collection of the admissible strategies that Player i goes last. More precisely, for $u_i = \{u_{in}\}_{n\in\mathbb{T}_t} \in \mathcal{U}_{t,i}(1)$, there exists a sequence of measurable functions $F_n(\cdot) \in U_i$ such that $u_{in} = F_n(\xi_k, k \le n; u_{1k}, u_{2k}, k < n)$ is the actual action taken by Player i at time n based upon his/her available information. Similarly, for $u_i = \{u_{in}\}_{n\in\mathbb{T}_t} \in \mathcal{U}_{t,i}(2)$, there exists a sequence of measurable functions $\tilde{F}_n(\cdot) \in U_i$ such that $u_{in} = \tilde{F}_n(\xi_k, k \le n; u_{1k}, u_{2k}, k < n; u_{jn},\ j \ne i)$.

Now, we are ready to define the upper and lower values,

$$V^+(t, x) = \inf_{u_1\in\mathcal{U}_{t,1}(1)} \sup_{u_2\in\mathcal{U}_{t,2}(2)} W(t, x, u_1, u_2) \tag{6}$$

and

$$V^-(t, x) = \sup_{u_2\in\mathcal{U}_{t,2}(1)} \inf_{u_1\in\mathcal{U}_{t,1}(2)} W(t, x, u_1, u_2), \tag{7}$$

respectively. If the lower value and upper value are equal to each other, then we say there exists a saddle point for the game, and its value is defined by

$$V(t, x) \stackrel{\Delta}{=} V^+(t, x) = V^-(t, x), \quad \forall (t, x) \in \mathbb{T}_0 \times S. \tag{8}$$

Our main interests are to identify the sufficient conditions for the existence of a saddle point.

3. Main result

In this part, we start with some basic properties of the upper and lower value functions in Section 3.1, and establish the sufficient conditions needed for the existence of a saddle point in Section 3.2.

3.1. Preliminary properties of the value function

First, we present a standard form of DPP:

Proposition 1. *Let V^+ and V^- be upper and lower values defined by (6) and (7). If (2) and (4) hold true, then we have following identities: for (t, x) satisfying $t < T$*

$$V^+(t, x) = \inf_{r_1\in U_1} \sup_{r_2\in U_2} \left\{e^{-\lambda_t} c_t(x, r_1, r_2) + \sum_{y\in S} p_t(x, y|r_1, r_2) V^+(t+1, y)\right\},$$

and

$$V^-(t, x) = \sup_{r_2\in U_2} \inf_{r_1\in U_1} \left\{e^{-\lambda_t} c_t(x, r_1, r_2) + \sum_{y\in S} p_t(x, y|r_1, r_2) V^-(t+1, y)\right\}.$$

Proof. By definition, we can write the cost function by