Operations Research Letters 46 (2018) 109–115

Contents lists available at ScienceDirect

Operations Research Letters

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

Continuous-time constrained stochastic games with average criteria

Wenzhao Zhang [a,*], Binfu Wang [a], Dewang Chen [a,b]

[a] College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, PR China
[b] Academy of Rail Transport, Fuzhou University, Fuzhou 350108, PR China

ARTICLE INFO

Article history:
Received 29 November 2016
Received in revised form 8 November 2017
Accepted 8 November 2017

Keywords:
Nonzero-sum game
Constrained Nash equilibria
Average criteria
Vanishing discount method

ABSTRACT

In this paper, we consider the continuous-time nonzero-sum stochastic games under the constrained average criteria. The state space is denumerable and the action space of each player is a general Polish space. The transition rates, reward and cost functions are allowed to be unbounded. The main hypotheses in this paper include the standard drift conditions, continuity-compactness condition and some ergodicity assumptions. By applying the vanishing discount method, we obtain the existence of stationary constrained average Nash equilibria.

1. Introduction

Constrained stochastic games constitute an important class of stochastic games with applications in many areas including queueing system, telecommunications systems, etc.; see, for instance, [2,14] and the reference therein. As far as our knowledge goes, the existing works for the nonzero-sum constrained stochastic games are mostly for the discrete-time case; see, for instance, [1,2,14]. More precisely, [1] studies the constrained stochastic games under the discounted and average cost criteria with finite state and finite action spaces. The constrained stochastic games with discounted cost criteria are also considered in [2] within the framework of countable state space and Borel action spaces. Under adequate conditions, [14] extends the result in [1] to the constrained average games with denumerable states. For continuous-time constrained game, the only work [15] considers the discounted cost criteria in denumerable state space.

In this paper, we study the continuous-time constrained game under the average criteria. The state space is denumerable but the action spaces are all Polish spaces. The transition rates, cost and reward functions may be unbounded *from above and from below*. Our main interest is to show the existence of constrained Nash equilibria out of the class of history-dependent strategy profiles. Roughly speaking, we want to obtain an equilibrium such that no single player can increase his own expected average reward by changing a strategy which satisfies constraints, while the other player continues to follow their original strategies. To this end, we use a refinement of classical assumptions and several technical results that have been introduced for the study of continuous-time Markov decision processes (CTMDPs) in [3,6,11]. First, we construct a sequence of auxiliary constrained discounted game models, in which the existence of constrained discounted Nash equilibria is proved by the new finite-state approximation technique in [3] under weaker condition than those in [15]. Next, we prove the limiting equilibria of constrained discounted Nash equilibria, as the discount factors tend to zero, are stationary constrained average Nash equilibria by means of the vanishing discount technique in [6,11]. The vanishing discount technique has been extensively applied in Markov decision processes (see [6,7,9,11,13]) and in stochastic games (see [5,14]).

The rest of the paper is organized as follows. In Section 2, we introduce the constrained average games we are concerned with. In Section 3, we present the auxiliary constrained discounted stochastic games. Moreover, we briefly recall some approximation properties of expected discounted and average reward/cost. We state these results only for the sake of completeness. In Section 4, we prove the existence of stationary constrained average Nash equilibria which can be viewed as the limits of stationary constrained Nash equilibria of the auxiliary constrained discounted game models. Some proofs of the auxiliary results are collected in Appendix A.

2. The game model

Notation. If X is a Borel space, we denote by $\mathcal{B}(X)$ its Borel σ-algebra and by $\mathbb{P}(X)$ the set of all probability measures on $\mathcal{B}(X)$ endowed with the topology of weak convergence. I stands for the indicator function, and $\delta_{\{x\}}(\cdot)$ is the Dirac measure concentrated at x. Define $\mathbb{R}_+ := (0, \infty)$ and $\mathbb{R}_+^0 := [0, \infty)$.

* Corresponding author.
E-mail addresses: zhangwenzhao@fzu.edu.cn, zhangwenzhao1987@163.com (W. Zhang).

https://doi.org/10.1016/j.orl.2017.11.006

Now, we introduce the following constrained stochastic game model $\mathcal{G}$:

$$\left\{S, A, B, \{A(i), B(i)|i\in S\}, r^1, r^2, (c^1,\theta^1), (c^2,\theta^2), q, \gamma\right\}, \tag{2.1}$$

where the state space $S := \{0,1,\ldots\}$ (i.e. the set of nonnegative integers) endowed with the discrete topology. The action spaces A for player 1 and B for player 2 are assumed to be Polish spaces endowed with the Borel σ-algebra $\mathcal{B}(A)$ and $\mathcal{B}(B)$, respectively. For each $i\in S$, $A(i)$ and $B(i)$ are the nonempty measurable subsets of A and B, respectively. Let $\mathbb{K} := \left\{(i,a,b)|i\in S, a\in A(i), b\in B(i)\right\}$, $K^1 := \left\{(i,a)|i\in S, a\in A(i)\right\}$, $K^2 := \left\{(i,b)|i\in S, b\in B(i)\right\}$. q denotes the transition rates satisfying that

- $q(j|i,a,b)\ge 0$ for all $(i,a,b)\in\mathbb{K}$ and $i\ne j$;
- $q(j|i,a,b)$ is measurable in $(a,b)\in A(i)\times B(i)$ for each fixed $i,j\in S$;
- $\sum_{j\in S} q(j|i,a,b) = 0$ for each $(i,a,b)\in\mathbb{K}$, and $q^*(i) := \sup_{a\in A(i),b\in B(i)} -q(i|i,a,b) < \infty$ for each $i\in S$.

For each $k=1,2$, the one-stage reward r^k and the one-stage costs c^k are real-valued functions on $\mathbb{K}$. The objective of each player k is to maximize the expected average reward corresponding to r^k, subject to the constraint that the expected average cost corresponding to c^k does not exceed the real number θ^k. Finally, γ denotes the initial distribution.

Next, we briefly recall the construction of the stochastic basis $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\ge 0}, P_\gamma^{\pi^1,\pi^2})$. Let $i_\infty\notin S$ be an isolated point and $S^* := S\bigcup\{i_\infty\}$. For each $n\ge 0$, $\Omega_n := S\times(\mathbb{R}_+\times S)^n\times(\{\infty\}\times\{i_\infty\})^\infty$, $\Omega := (S\times(\mathbb{R}_+\times S)^\infty)\cup\bigcup_{n=0}^\infty \Omega_n$. Thus, we obtain the sample space $(\Omega,\mathcal{F})$, where $\mathcal{F}$ is the standard Borel σ-algebra. For each $m\ge 1$ and each sample $\omega=(i_0,\theta_1,i_1,\ldots,\theta_{n-1},i_{n-1},\ldots)\in\Omega$, we define some maps on Ω as follows: $T_0(\omega):=0$, $X_0(\omega):=i_0$, $\Theta_m(\omega):=\theta_m$, $T_m(\omega) := \sum_{n=1}^m\theta_n$, $T_\infty(\omega) := \lim_{n\to\infty}T_m(\omega)$, $X_m(\omega) := i_m$. Here, Θ_m, T_m, X_m denote the sojourn time, jump moment and the state of the process on the interval $[T_m, T_{m+1})$, respectively. Let us define a process $\{\xi_t, t\ge 0\}$ on $(\Omega,\mathcal{F})$ by

$$\xi_t(\omega) := \sum_{m\ge 0} I_{\{T_m(\omega)\le t<T_{m+1}(\omega)\}} i_m + I_{\{T_\infty(\omega)\le t\}} i_\infty \text{ for each } \omega\in\Omega.$$

In what follows, $h_n(\omega) := (i_0,\theta_1,i_1,\ldots,\theta_n,i_n)$ is the n-component internal history, the argument $\omega=(i_0,\theta_1,i_1,\ldots,\theta_n,i_n,\ldots)\in\Omega$ is often omitted. Since we do not plan to consider the process after T_∞, i_∞ is regarded as absorbing. Let us define $A(i_\infty) := a_\infty$ and $B(i_\infty) := b_\infty$, where $a_\infty\notin A$ and $b_\infty\notin B$ are two isolated points, $A^* := A\cup\{a_\infty\}$, $B^* := B\cup\{b_\infty\}$ and $q(i_\infty|i_\infty,a_\infty,b_\infty) := 0$. Let $\mathcal{F}_t := \sigma(\{T_m\le s, X_m=j\} : j\in S, s\le t, m\ge 0)$ be the internal history to time t for the game model $\mathcal{G}$, $\mathcal{F}_{s-} := \bigvee_{t<s}\mathcal{F}_t$, $\mathcal{P} := \sigma(C\times\{0\}(C\in\mathcal{F}_0), C\times(s,\infty)(C\in\mathcal{F}_{s-}, s>0))$ which denotes the predictable σ-algebra on $\Omega\times\mathbb{R}_+^0$.

Below we introduce the concept of strategies.

Definition 2.1.

(i) A $\mathcal{P}$-measurable transition probability function $\pi^1(\cdot|\omega,t)$ on $(A^*,\mathcal{B}(A^*))$, concentrated on $A(\xi_{t-}(\omega))$, is called a *randomized history-dependent strategy* for player 1.

(ii) A randomized history-dependent strategy π^1 for player 1 is said to be *randomized stationary* if there exists a stochastic kernel φ^1 on A^* given S^* such that $\pi^1(\cdot|\omega,t) = \varphi^1(\cdot|\xi_{t-}(\omega))$. Such strategies are denoted as φ^1.

With set B^* in lieu of set A^*, we define similarly the randomized history-dependent strategy and randomized stationary strategy for player 2, denoted by π^2 and φ^2, respectively. We denote by Π^k the class of randomized history-dependent strategies for player k, and by Φ^k the class of randomized stationary strategies for player k. For each strategy profile $\boldsymbol{\pi} := (\pi^1,\pi^2)\in\Pi^1\times\Pi^2$, and strategy $\pi'^k\in\Pi^k$, we denote by $[\boldsymbol{\pi}^{-k},\pi'^k]$ the *strategy profile* obtained from $\boldsymbol{\pi}$ by replacing π^k with π'^k.

Remark 2.1. Unlike most of works on continuous-time stochastic game restricted in the set of all Markov strategies, we take history-dependent strategies into consideration. However, we do not assume that the players choose their action based on the observation of previous actions here. Proposition 1 of [10] shows that the strategy of each player should be independent on his/her own previous actions, as otherwise, there are strategies that do not define a play. Meanwhile, we assume here that each player cannot observe the other player's actions, which is frequently used in game theory.

Now, let us briefly recall the construction of $P_\gamma^{\pi^1,\pi^2}$. Let $H_0 := S$ and $H_m := S\times((0,\infty]\times S^*)^m$ be the set of histories for each $m\ge 1$. For each strategy profile $(\pi^1,\pi^2)\in\Pi^1\times\Pi^2$, according to Theorem 4.19 in [8], (π^1,π^2) has the following form:

$$\pi^1(da|\omega,t) = \delta_{\{a_\infty\}}(da)I_{\{t\ge T_\infty\}} + \sum_{m=0}^\infty I_{\{T_m<t\le T_{m+1}\}} \times \pi_m^1(da|i_0,\theta_1,\ldots,i_m,t-T_m),$$

$$\pi^2(db|\omega,t) = \delta_{\{b_\infty\}}(db)I_{\{t\ge T_\infty\}} + \sum_{m=0}^\infty I_{\{T_m<t\le T_{m+1}\}} \times \pi_m^2(db|i_0,\theta_1,\ldots,i_m,t-T_m),$$

where π_m^1 (resp., π_m^2) is some stochastic kernel on A (resp., B) given $H_m\times\mathbb{R}_+$ for each $m\ge 0$. Then, for each $m\ge 0$, $t\ge 0$ and $h_m\in H_m$, let us define

$$\Lambda^{m,\pi^1,\pi^2}(j|h_m,t) := \int_A \pi_m^1(da|h_m,t)\int_B \pi_m^2(db|h_m,t) \times q(j\setminus\{i_m\}|i_m,a,b),$$

where $q(j\setminus\{i_m\}|i_m,a,b) := q(j|i_m,a,b)I_{\{j\ne i_m\}}$, and $\Lambda^{m,\pi^1,\pi^2}(S|h_m,t) := \sum_{j\in S}\Lambda^{m,\pi^1,\pi^2}(j|h_m,t)$. The marginal of $P_\gamma^{\pi^1,\pi^2}$ on H_0 coincides with γ. Suppose that the measure $P_\gamma^{\pi^1,\pi^2}$ on H_m has been constructed, the marginal of $P_\gamma^{\pi^1,\pi^2}$ on H_{m+1} is defined by

$$P_\gamma^{\pi^1,\pi^2}(\Gamma^{H_m}\times(dt,j)) := \int_{\Gamma^{H_m}} P_\gamma^{\pi^1,\pi^2}(dh_m)I_{\{X_m\in S\}}\Lambda^{m,\pi^1,\pi^2}(j|h_m,t) \times e^{-\int_0^t \Lambda^{m,\pi^1,\pi^2}(S|h_m,v)dv}dt$$

$$P_\gamma^{\pi^1,\pi^2}(\Gamma^{H_m}\times(\infty,i_\infty)) := \int_{\Gamma^{H_m}} P_\gamma^{\pi^1,\pi^2}(dh_m)\Big\{I_{\{X_m=i_\infty\}} + I_{\{X_m\in S\}} \times e^{-\int_0^\infty \Lambda^{m,\pi^1,\pi^2}(S|h_m,v)dv}\Big\}, \tag{2.2}$$

for each $\Gamma^{H_m}\in\mathcal{B}(H_m)$. According to the well known Tulcea's Theorem, there exists a unique probability measure $P_\gamma^{\pi^1,\pi^2}$ on $(\Omega,\mathcal{F})$ such that its marginal onto H_m satisfies (2.2). Hence, we get a stochastic basis $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\ge 0},P_\gamma^{\pi^1,\pi^2})$, which is always assumed to be complete. Expectations with respect to $P_\gamma^{\pi^1,\pi^2}$ is denoted as $E_\gamma^{\pi^1,\pi^2}$. When $\gamma(i)=1$, we write $P_i^{\pi^1,\pi^2}$ for $P_\gamma^{\pi^1,\pi^2}$ and $E_i^{\pi^1,\pi^2}$ for $E_\gamma^{\pi^1,\pi^2}$, respectively.

Roughly speaking, the game is played as follows. At each time t, the players can only observe the current state, and the past states and jump moments. If $T_n<t\le T_{n+1}$ for some n, both players independently and simultaneously choose actions $a\in A(i_n)$ and $b\in B(i_n)$ according to $\pi_n^1(\cdot|h_n,t-T_n)$ and $\pi_n^2(\cdot|h_n,t-T_n)$, respectively. Then, the following happen: (1) player k $(k=1,2)$ has reward and cost at the rate $\int_A\int_B r^k(i_n,a,b)\pi_n^1(da|h_n,t-T_n)\pi_n^2(db|h_n,t-T_n)$ and