KYBERNETIKA — VOLUME *58* (2022), NUMBER 2, PAGES 163–179

DETERMINISTIC MARKOV NASH EQUILIBRIA FOR POTENTIAL DISCRETE-TIME STOCHASTIC GAMES

Alejandra Fonseca-Morales

In this paper, we study the problem of finding deterministic (also known as feedback or closed-loop) Markov Nash equilibria for a class of discrete-time stochastic games. In order to establish our results, we develop a potential game approach based on the dynamic programming technique. The identified potential stochastic games have Borel state and action spaces and possibly unbounded nondifferentiable cost-per-stage functions. In particular, the team (or coordination) stochastic games and the stochastic games with an action independent transition law are covered.

Keywords: stochastic games, optimal control, potential approach, dynamic programming

Classification: 91A50, 91A25, 93E20, 91A14, 91A10

1. INTRODUCTION

We study noncooperative discrete-time stochastic games with Borel action and state spaces and possibly unbounded nondifferentiable cost-per-stage functions. Our main objective is to identify some classes of potential stochastic games in this family. Potential stochastic games are stochastic games for which we can associate an optimal control problem (OCP) whose (Markov) optimal solutions are (Markov) Nash equilibria for the concerned game. Note that, the problem of finding Nash equilibria is really simplified because instead of solving N *coupled* OCPs, as in the standard case, one can solve a *single system*, which corresponds to the associated OCP. Besides, we obtain *deterministic* Markov Nash equilibria as opposed to mixed (or randomized) equilibria. Moreover, by applying the dynamic programming technique, we can prove that several properties from control theory could be held for that Nash equilibria obtained by solving an OCP.

Potential games have been studied under several scenarios. For instance, for static games see [13, 16, 17] and for dynamic games see [1, 3, 4, 5, 6, 10], [14], [19], and [20]. See also [11] and [18] which have not yet been published. In particular, works [5, 10, 11, 14] address problems related to the results of our paper but under more restrictive conditions. Indeed, [5] considers stochastic games with Borel action and state spaces, differentiable functions and a discrete-time equation which describes the state trayectory under an uncoupled condition among the players. [10] identifies potential

DOI: 10.14736/kyb-2022-2-0163

deterministic dynamic games, i.e, dynamic games where the state dynamics is deterministic. [11] establishes potential stochastic games with finite state space and Lyapunov conditions for the payoff functions. Finally, [14] presents dynamic stochastic games with separable reward functions, an action independent transition law and a finite state space. Therefore, to the best of our knowledge, our results generalize these works because we identify potential stochastic games with Borel action and state spaces and possibly unbounded and nondifferentiable cost-per-stage functions. Furthermore, our results include team (or coordination) dynamic stochastic games and dynamic stochastic games with an action independent transition law.

The paper is organized as follows. In Sections 2 and 3, we introduce some models of discrete-time stochastic games and stochastic OCP models that we are interested in, respectively. We present our main results assuming a finite-horizon in Sections 4 and 5. The infinite-horizon case is done in Section 6. Finally in Section 7, we give some conclusions.

2. DISCRETE-TIME STOCHASTIC GAMES

In this section, we introduce the game model and some basic definitions.

We consider the discrete-time stochastic game model

$$(\bar{N}, T, X, A_i, \{A_i(x)\}, Q, r^i, c_T) \tag{1}$$

whose components are specified as follows.

Let $\bar{N} = \{1, \ldots, N\}, N \geq 2$, be the set of players and $T < \infty$ the horizon. The sets X and A_i are Borel spaces denoting the *state* space and the *action* or *control* set for the player $i \in \bar{N}$, respectively. For each state $x \in X$, the feasible action sets $A_i(x) \subset A_i$ are nonempty measurable sets, and the set of state-action vectors is defined by

$$\mathbb{K} := \Big\{(x, a^1, \ldots, a^N) | x \in X, a^1 \in A_1(x), \ldots, a^N \in A_N(x)\Big\}.$$

The transition law among states is a stochastic kernel on X given $\mathbb{K}$, represented by

$$Q(B|x, a^1, \ldots, a^N), \;\; B \in \mathscr{B}(X),$$

where $\mathscr{B}(X)$ is the Borel $\sigma-$algebra on X.

The function $r^i : \mathbb{K} \to \mathbb{R}$ is the cost-per-stage for player i, while $c_T : X \to \mathbb{R}$ is a final cost function.

Let $\pi^i = \{\pi^i_t\}$ be a sequence of stochastic kernels π^i_t on A_i given $\mathbb{K}^t \times X$. Hence, the set of (mixed) strategies for player i is

$$\Pi_i := \Big\{\pi^i = \{\pi^i_t\}_{t=0}^{T-1} \;\mid\; \pi^i_t(A_i(x_t)|h_t) = 1 \;\; \forall \;\; h_t \in \mathbb{K}^t \times X\Big\},$$

where $h_t \in \mathbb{K}^t \times X$ represents the history of the game up to time t. Thus, the space of multi-strategies for the players is $\Pi = \Pi_1 \times \cdots \times \Pi_N$.

We shall restrict to the family of Markov policies, this is, π^i_t depens only on the current state x_t and not on the full $t-$ history h_t.

Now, let $\mathbb{F}_i$ be the set of measurable selectors for player i, that is,

$$\mathbb{F}_i := \Big\{ f^i : X \to A_i \;\mid\; f^i(x) \in A_i(x) \;\; \forall x \in X \Big\}.$$

We define $\mathbb{F} := \mathbb{F}_1 \times \cdots \times \mathbb{F}_N$ and assume that $\mathbb{F} \neq \emptyset$.

Definition 2.1. A *deterministic (or pure) Markov multi-strategy* is a sequence $f = \{f_t\}_{t=0}^{T-1}$ of functions such that $f_t \in \mathbb{F}$.

The class of deterministic Markov multi-strategies is also known in the literature as the feedback or closed-loop multi-strategy space. In contrast, an open-loop strategy is formed by selectors depending only on the initial condition and time.

For each $\pi \in \Pi$ and initial state $x \in X$, we define the expected total cost $U_i : X \times \Pi \to \mathbb{R}$, for the player i as

$$U_i(x,\pi) \quad := \quad E_x^{\pi}\left[\sum_{t=0}^{T-1} r^i(x_t,a_t) + c_T(x_T)\right]. \tag{2}$$

Note that the final cost c_T is the same for all players.

We recall the usual notation for multi-strategies

$$\begin{aligned}(a^i, a^{-i*}) &:= (a^{1*},\ldots,a^{i-1*},a^i,a^{i+1*},\ldots,a^{N*}),\\ (\pi^i, \pi^{-i*}) &:= (\pi^{1*},\ldots,\pi^{i-1*},\pi^i,\pi^{i+1*},\ldots,\pi^{N*}).\end{aligned}$$

Definition 2.2. A Nash equilibrium is a multi-strategy $\pi^* \in \Pi$ that satisfies, for any $x \in X$ and $i \in \bar{N}$,

$$U_i(x,\pi^*) \quad \le \quad U_i(x,\pi^i,\pi^{-i*}) \;\; \forall \pi^i \in \Pi_i.$$

Likewise, a deterministic Markov multi-strategy $f^* = \{f_t^*\}_{t=0}^{T-1} \in \mathbb{F}$ is a (Markov) Nash equilibrium if

$$U_i(x,f^*) \quad \le \quad U_i(x,\pi^i,f^{-i*}) \;\; \forall \pi^i \in \Pi_i.$$

Now, considering a dynamic game, as in (1)−(2), we introduce the following definition:

Definition 2.3. A (closed-loop) potential dynamic game is a dynamic game such that it is possible to associate to it an OCP whose (closed-loop) optimal solutions are also (closed-loop) Nash equilibria for the dynamic game.

Hence, finding an appropiate OCP in the sense of Definition 2.3 raises the problem of finding a functional U, which satisfies, for each $x \in X$ and $\pi^* \in \Pi, \pi^i \in \Pi_i$,

$$U(x,\pi^*) - U(x,\pi^i,\pi^{-i*}) = U_i(x,\pi^*) - U_i(x,\pi^i,\pi^{-i*}). \tag{3}$$

Furthermore, as in the case when the game problem is restricted to open-loop strategies, we prove below that it is possible to obtain U by finding a function $P : \mathbb{K} \to \mathbb{R}$, called a potential function, such that for each $x \in X$, $a^* \in A(x) := \prod_{i=1}^N A_i(x)$, and $a^i \in A_i(x)$,

$$P(x, a^*) - P(x, a^i, a^{-i*}) = r^i(x, a^*) - r^i(x, a^i, a^{-i*}). \tag{4}$$

Although additional conditions will be also asked for the dynamics of the game.

It is worth remarking that the advantage of this potential approach is to simplify the game problem finding a suitable OCP. Hence, the game problem becomes in a standard minimization problem, and therefore we can obtain a Nash equilibrium in the set of deterministic Markov multi-strategies.

3. OPTIMAL CONTROL PROBLEMS

We define an optimal control problem taking some components from the game (1) – (2) together with $P : X \times A \to \mathbb{R}$ as the stage-cost function with $A := \prod_{i=1}^N A_i$.

Consider the control model

$$(T, X, A, \{A(x) = \prod_{i=1}^{N} A_i(x)\}, Q, P, c_T), \tag{5}$$

with the performance index $U : X \times \Pi \to \mathbb{R}$ given by

$$U(x, \pi) \quad := \quad E_x^\pi \left[\sum_{t=0}^{T-1} P(x_t, a_t) + c_T(x_T) \right]. \tag{6}$$

In this case, the value function is

$$U^*(x) := \inf_{\pi \in \Pi} U(x, \pi), \quad x \in X. \tag{7}$$

Definition 3.1. An optimal solution for the OCP (5) – (7) is a strategy $\pi^* \in \Pi$ such that

$$U^*(x) = U(x, \pi^*) \quad \forall x \in X$$

As is well-known in the literature of Markov decision processes, the optimal strategy $\pi^* \in \Pi$ in Definition 3.1 can be found in the set of deterministic Markov strategies by using dynamic programming arguments. This fact is stated in the following result, see Sections 3.2 and 3.3 in [7].

Theorem 3.2. Consider an OCP as in (5) – (7). If $\pi^* \in \Pi$ is an optimal solution, then there exists a deterministic Markov strategy $f^* = \{f_t^*\}_{t=0}^{T-1}$ such that, for every $x \in X$,

$$\begin{aligned} J_T(x) &:= c_T(x), \\ J_t(x) &:= \min_{\prod_{i=1}^N A_i(x)} \left[P(x, a) + \int_X J_{t+1}(y) Q(\mathrm{d}y | x, a) \right] \\ &= P(x, f_t^*(x)) + \int_X J_{t+1}(y) Q(\mathrm{d}y | x, f_t^*(x)), \end{aligned} \tag{8}$$

and $U^*(x) = U(x, \pi^*) = J_0(x) = U(x, f^*)$.

4. THE POTENTIAL APPROACH

To develop the potential approach, we requiere to assume the existence of a function P satisfying (4) and a solution of the corresponding OCP. Regarding the latter, there are several conditions that guarantee the existence of such solutions, see [7, 8]. However, for our purposes, we are going to take this scenario for granted which we establish in the following assumptions.

Assumption 4.1. There is a function P satisfying (4).

Assumption 4.2. The OCP (5) – (7) has an optimal solution π^*. That is, there is a deterministic Markov optimal solution $f^* = \{f_t^*\}_{t=0}^{T-1}$, where each $f_t^* = (f_t^{1*}, \dots, f_t^{N*}) \in \mathbb{F}$ satisfies (8).

Define for each deterministic Markov multi-strategy $f = \{f_t\}_{t=0}^{T-1}$, and each time $t = 1, \dots, T-1$ and $x \in X$,

$$P(t, x, f) := E_x^f \sum_{n=t}^{T-1} P(x_n, a_n) \qquad \text{and} \qquad r^i(t, x, f) := E_x^f \sum_{n=t}^{T-1} r^i(x_n, a_n).$$

Assumption 4.3. For each $t = 0, \dots, T-2, x \in X, i \in \bar{N}, a^i \in A_i$, and a deterministic Markov multi-strategy $f = \{f_t\}_{t=0}^{T-1}$, it holds

$$\begin{aligned} &\int_X P(t+1, y, f) Q(\mathrm{d}y|x, f_t(x)) - \int_X P(t+1, y, f) Q(\mathrm{d}y|x, a^i, f_t^{-i}(x)) = \\ &\quad \int_X r^i(t+1, y, f) Q(\mathrm{d}y|x, f_t(x)) - \int_X r^i(t+1, y, f) Q(\mathrm{d}y|x, a^i, f_t^{-i}(x)). \end{aligned} \tag{9}$$

For Assumption 4.3 to make sense, we shall assume that P and r^i are almost everywhere integrable functions with respect to Q. Moreover, Assumption 4.3 is a similar condition to (4) but on the transition kernel Q.

In the following lemma we establish a relationship between the solutions of a dynamic game (1) – (2) and that of the OCP (5) – (7). To this end, for each $i = 1, \dots, N$, and $f^* = \{f_t^*\}_{t=0}^{T-1}$ as in Assumption 4.2, let $J_0^i, \dots, J_T^i$ be defined by

$$\begin{aligned} J_T^i(x) &:= c_T(x), \\ J_t^i(x) &:= \min_{a^i \in A_i(x)} \left[r^i(x, a^i, f_t^{-i*}(x)) + \int_X J_{t+1}^i(y) Q(\mathrm{d}y|x, a^i, f_t^{-i*}(x)) \right]. \end{aligned} \tag{10}$$

To ease notation we will write f_t^* instead of $f_t^*(x)$.

For the remainder of this section, we assume that Assumptions 4.1-4.3 hold true.

Lemma 4.4. For the dynamic game (1) – (2), each i–th component $f^{i*} = \{f_t^{i*}\}_{t=0}^{T-1}$ of an optimal deterministic Markov strategy $f^* = \{f_t^*\}_{t=0}^{T-1}$ for the OCP (5) – (7) defined by P, attains the minimum in (10) and satisfies

$$U_i(x, f^{i*}, f^{-i*}) = J_0^i(x) = \inf_{\pi^i \in \Pi_i} U_i(x, \pi^i, f^{-i*}).$$

P r o o f. Take an arbitrary player i and fix the multi-strategy f^{-i*} corresponding to the optimal solution f^*. Define the following functions, for each $x \in X$ and $\pi^i \in \Pi_i$,

$$\begin{aligned} C^i_T(x) &:= c_T(x), \\ C^i_t(x, \pi^i, f^{-i*}) &:= E_x^{\pi^i, f^{-i*}} \left[\sum_{n=t}^{T-1} r^i(x_n, a^i_n, a^{-i}_n) + c_T(x_T) \right]. \end{aligned}$$

We shall show by backward induction over $t = 0, \ldots, T$ that for player i, the following hold:

(a) $C^i_t(x, \pi^i, f^{-i*}) \geq J^i_t(x)$ for all $\pi^i \in \Pi_i$ and,

(b) $C^i_t(x, f^{i*}, f^{-i*}) = J^i_t(x)$.

To prove (a), note that $C^i_T(x) = c_T(x) = J^i_T(x)$, and

$$\begin{aligned} C^i_{T-1}(x, \pi^i, f^{-i*}) &= \int_{A_i} \left[r^i(x, a^i, f^{-i*}_{T-1}) + \int_X J^i_T(y) Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-1}) \right] \pi^i_{T-1}(da^i|x) \\ &\geq J^i_{T-1}(x). \end{aligned}$$

To use an induction argument, assume now that $C^i_{t+1}(x, \pi^i, f^{-i*}) \geq J^i_{t+1}(x)$. Therefore, for t, we have

$$\begin{aligned} C^i_t(x, \pi^i, f^{-i*}) &\geq \int_{A_i} \left[r^i(x, a^i, f^{-i*}_t) + \int_X J^i_{t+1}(y) Q(\mathrm{d}y|x, a^i, f^{-i*}_t) \right] \pi^i_t(da^i|x) \\ &\geq J^i_t(x). \end{aligned}$$

Hence, (a) follows.

Now, for (b), notice that for $T-1$, by Dynamic Programming arguments give that

$$\begin{aligned} P(x, f^{i*}_{T-1}, f^{-i*}_{T-1}) &- P(x, a^i, f^{-i*}_{T-1}) \\ + \int_X c_T(y) Q(\mathrm{d}y|x, f^{i*}_{T-1}, f^{-i*}_{T-1}) &- \int_X c_T(y) Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-1}) \leq 0, \end{aligned}$$

for all $a^i \in A_i$. Thus, by (4)

$$\begin{aligned} C^i_{T-1}(x, f^{i*}, f^{-i*}) &= r^i(x, f^{i*}_{T-1}, f^{-i*}_{T-1}) + \int_X c_T(y) Q(\mathrm{d}y|x, f^{i*}_{T-1}, f^{-i*}_{T-1}) \\ &\leq r^i(x, a^i, f^{-i*}_{T-1}) + \int_X c_T(y) Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-1}) \ \ \forall a^i \in A_i. \end{aligned}$$

Therefore, $C^i_{T-1}(x, f^{i*}, f^{-i*}) = J^i_{T-1}(x)$.

Similarly, for $T-2$ and any $a^i \in A_i$, we have,

$$\begin{aligned} P(x, f^*_{T-2}) &- P(x, a^i, f^{-i*}_{T-2}) \\ + \int_X P(y, f^*_{T-1}) Q(\mathrm{d}y|x, f^*_{T-2}) &- \int_X P(y, f^*_{T-1}) Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-2}) \\ + \int_X E^{f^*}_y[c_T(x_T)] Q(\mathrm{d}y|x, f^*_{T-2}) &- \int_X E^{f^*}_y[c_T(x_T)] Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-2}) \leq 0. \end{aligned}$$

Then, by (4) and Assumption 4.3

$$\begin{aligned} C^i_{T-2}(x, f^{i*}, f^{-i*}) &= r^i(x, f^*_{T-2}) + \int_X r^i(y, f^*_{T-1}) Q(\mathrm{d}y|x, f^*_{T-2}) \\ &\quad + \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, f^*_{T-2}) \\ &\leq r^i(x, a^i, f^{-i*}_{T-2}) + \int_X r^i(y, f^*_{T-1}) Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-2}) \\ &\quad + \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, a^i, f^{-i*}_{T-2}), \end{aligned}$$

for all $a^i \in A_i$. Thus, $C^i_{T-2}(x, f^{i*}, f^{-i*}) = J^i_{T-2}(x)$.

Assume that $C^i_{t+1}(x, f^{i*}, f^{-i*}) = J^i_{t+1}(x)$. Now for t, we have that for any $a^i \in A_i$,

$$\begin{aligned} & P(x, f^*_t) - P(x, a^i, f^{-i*}_t) \\ &+ \int_X P(t+1, y, f^*) Q(\mathrm{d}y|x, f^*_t) - \int_X P(t+1, y, f^*) Q(\mathrm{d}y|x, a^i, f^{-i*}_t) \\ &+ \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, f^*_t) - \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, a^i, f^{-i*}_t) \leq 0. \end{aligned}$$

Thus, the previous equation implies that, for any $a^i \in A_i$,

$$\begin{aligned} C^i_t(x, f^{i*}, f^{-i*}) &= r^i(x, f^*_t) + \int_X r^i(t+1, y, f^*) Q(\mathrm{d}y|x, f^*_t) \\ &\quad + \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, f^*_t) \\ &\leq r^i(x, a^i, f^{-i*}_t) + \int_X r^i(t+1, y, f^*) Q(\mathrm{d}y|x, a^i, f^{-i*}_t) \\ &\quad + \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, a^i, f^{-i*}_t) \\ &= r^i(x, a^i, f^{-i*}_t) + \int_X J^i_{t+1}(y) Q(\mathrm{d}y|x, a^i, f^{-i*}_t) \\ &\quad + \int_X E^{f^*}_y [c_T(x_T)] Q(\mathrm{d}y|x, a^i, f^{-i*}_t), \end{aligned}$$

which proves (b). □

Theorem 4.5. The dynamic game (1) – (2) is a closed-loop potential dynamic game.

P r o o f. We consider the OCP (5) – (7) to find a solution for the game (1) – (2). That is, if $f^* = \{f^*_t\}^{T-1}_{t=0}$ is an optimal solution for the corresponding OCP, then by Lemma 4.4, for every $i \in N$,

$$U_i(x, f^*) = J^i_0(x) \quad \leq \quad C^i_0(x, \pi^i, f^{-i*}) = U_i(x, \pi^i, f^{-i*}) \;\; \forall \pi^i \in \Pi_i.$$

In other words, f^* is a deterministic Nash equilibrium for the dynamic game (1) – (2). □

5. CLOSED-LOOP POTENTIAL STOCHASTIC GAMES

In this section, we focus on the study of the previously imposed assumptions. We will specifically introduce particular classes of stochastic dynamic games that satisfy Assumptions 4.1 and 4.3.

Consider dynamic games, as in (1) – (2), where the functions r^i are of the form

$$r^i(x,a) \quad := \quad P(x,a) + L^i(x, a^1, \ldots, a^{i-1}, a^{i+1}, \ldots, a^N), \tag{11}$$

where $P : X \times A \to \mathbb{R}$ and $L^i : X \times A_{-i} \to \mathbb{R}$ are known measurable functions and $a = (a^1, \ldots, a^{i-1}, a^i, a^{i+1}, \ldots, a^N)$.

Note that the function P in (11) satisfies (4) because, for each i and $a^i \in A_i$,

$$\begin{aligned} r^i(x,a^*) - r^i(x,a^i,a^*) \quad &:= \quad P(x,a^*) + L^i(x, a^{1*}, \ldots, a^{i-1*}, a^{i+1*}, \ldots, a^{N*}) \\ &- \quad [P(x,a^i,a^*) + L^i(x, a^{1*}, \ldots, a^{i-1*}, a^{i+1*}, \ldots, a^{N*})] \\ &= \quad P(x,a^*) - P(x,a^i,a^*). \end{aligned}$$

In other words, the game (1) – (2) with cost-per-stage as in (11) satisfies Assumption 4.1.

Definition 5.1. We say that a dynamic game, as in (1) – (2), has an action-independent transition law, if Q satisfies, for each $(x,a) \in \mathbb{K}$,

$$Q(\cdot|x,a) = Q(\cdot|x).$$

Note that an action-independent transition law trivially satisfies Assumption 4.3. For illustration, see Examples 5.15 and 6.5, below.

Proposition 5.2. The dynamic game (1) – (2) under Assumption 4.2 with functions r^i and P as in (11) and an action independent transition law Q, is a closed-loop potential game, which has a deterministic Markov Nash equilibrium.

Proof. As we have shown at the begining of the section, this result follows directly from Theorem 4.5 given that (4) and Assumption 4.3 hold. □

Similar results are given in [5] and [14] where an action independent transition law is assumed. In [5] it is assumed an uncoupled condition among several states, while [14] considers a finite state space.

Assumption 5.3. For each deterministic Markov multi-strategy $f : X \longrightarrow A(x)$, the functions L^i in (11) satisfy

$$L^i(x, f^{-i}(x)) = L^i(y, f^{-i}(y)) \;\; \forall y \in X.$$

In other words, $L^i_{f^{-i}} := L^i(x, f^{-i}(x))$ is constant in the variable x.

Proposition 5.4. Under Assumptions 4.2 and 5.3, the dynamic game (1) – (2) with functions r^i and P as in (11) is a closed-loop potential game with a deterministic Markov Nash equilibrium.

P r o o f. Note that Assumption 5.3 implies Assumption 4.3 because for all $a^i \in A_i$,

$$\int_X E_y^f \sum_{n=t+1}^{T-1} L^i(x_n, a_n^{-i}) Q(\mathrm{d}y|x, a^i, f^{-i}) = \sum_{n=t+1}^{T-1} L^i_{f_n^{-i}}.$$

Thus, by Theorem 4.5, the result follows.

□

It is worth noting the difference between Propositions 5.2 and 5.4. In Proposition 5.2, no restritions are being imposed on the functions L^i, while in Proposition 5.4, restrictions are not imposed on the transition kernel Q.

A version of Assumption 5.3 is presented in [11] where the author assumes a finite state space.

According to Proposition 5.4, we can state the following result.

Corollary 5.5. The dynamic game (1) – (2) with functions r^i and P as in (11) is a closed-loop potential dynamic game when the functions $L^i, i \in \bar{N}$ are constants, namely, $L^i = k_i$.

Note that Corollary 5.5 includes the classical team (or coordination) games, that is, a game with $r^i(\cdot) = r(\cdot)$ for all $i \in \bar{N}$ to be closed-loop potential games, a team dynamic game is a game as is described in Corollary 5.5 but letting $k_i \equiv 0$ for all $i \in \bar{N}$. In other words, a game where all the players have the same payoff function P. Hence, our results are an extension of those in [14] where team dynamic stochastic games with a finite state space are studied via a similar potential approach.

Following the previous ideas, the next condition is also sufficient for Assumption 4.3.

Assumption 5.6. For each $x \in X$ and $f, g \in \mathbb{F}$, the functions L^i and the transition kernel Q satisfy

$$\int_X L^i(y, f^{-i}(y)) Q(\mathrm{d}y|x, g(x)) = L^i(x, f^{-i}(x)).$$

Note that Assumption 5.6 generalizes Assumption 5.3.

Proposition 5.7. Under Assumptions 4.2 and 5.6 the dynamic game (1) – (2) with functions r^i and P as in (11) is a closed-loop potential dynamic game with a deterministic Markov Nash equilibrium.

P r o o f. Assumption 5.6 implies Assumption 4.3 because for each $x \in X$,

$$\int_X E_y^{f^i, f^{-i}} \sum_{n=t+1}^{T-1} L^i(x_n, a_n^{-i}) Q(\mathrm{d}y|x, a^i, f_t^{-i}) \quad = \quad \sum_{n=t+1}^{T-1} L^i(x, f_n^{-i}(x)),$$

for every $a^i \in A_i(x)$. □

Observe that Assumption 4.3 requires that (9) holds for all deterministic Markov multi-strategy. Instead, in the following Assumption 5.8 we only require to impose conditions using an optimal strategy satisfying Assumption 4.2.

Assumption 5.8. The transition law Q has a density function $q : \mathbb{K} \times X \to \mathbb{R}$ with respect to a measure λ, i. e.,

$$Q(B|x,a) := \int_B q(x,a,y)\lambda(\mathrm{d}y).$$

Moreover, the functions $L^i \geq 0$ and q satisfies, for each i,

$$q(x_t, a^i, f_t^{-i*}(x), y) - q(x_t, f_t^*(x), y) \geq 0 \;\; \forall a^i \in A_i(x_t), t = T-1, \ldots, 0,$$

where f_t^* is as in Assumption 4.2.

Proposition 5.9. The dynamic game (1) – (2) with functions r^i, P as in (11) and Q as in Assumption 5.8 is a closed-loop potential dynamic game with a deterministic Markov Nash equilibrium.

Proof. It follows using similar ideas as Lemma 4.4. Note that (a) in the proof of Lemma 4.4 is true. We only require to show (b), that is, $C_t^i(x, f^{i*}, f^{-i*}) = J_t^i(x)$. For each $a^i \in A_i$, and $t = 1, \ldots, T-1$,

$$\int_X E_y^{f^*} \sum_{n=t+1}^{T-1} L^i(x, a_n^{-i})[q(x, a^i, f_t^{-i*}, y) - q(x, f_t^*, y)]\lambda(\mathrm{d}y) \quad \geq \quad 0.$$

In other words,

$$\begin{aligned}
C_t^i(x, f^{i*}, f^{-i*}) \quad = \quad & P(x, f_t^*) + L^i(x, f_t^{-i*}) \\
+ \quad & \int_X E_y^{f^*} \sum_{n=t+1}^{T-1} P(x_n, a_n) q(x, f_t^*, y)\lambda(\mathrm{d}y) \\
+ \quad & \int_X E_y^{f^*} \sum_{n=t+1}^{T-1} L^i(x, a_n^{-i}) q(x, f_t^*, y)\lambda(\mathrm{d}y) \\
+ \quad & \int_X E_y^{f^*}[c_T(x_T)] Q(\mathrm{d}y|x, f_t^*) \\
= \quad & \min_{a^i \in A_i(x)} [P(x, a^i, f_t^{-i*}) + L^i(x, f_t^{-i*}) \\
+ \quad & \int_X E_y^{f^*} \sum_{n=t+1}^{T-1} P(x, a_n) q(x, a^i, f_t^{-i*}, y)\lambda(\mathrm{d}y) \\
+ \quad & \int_X E_y^{f^*} \sum_{n=t+1}^{T-1} L^i(x, a_n^{-i}) q(x, a^i, f_t^{-i*}, y)\lambda(\mathrm{d}y) \\
+ \quad & \int_X E_y^{f^*}[c_T(x_T)] Q(\mathrm{d}y|x, a^i, f_t^{-i*})] \\
= \quad & J_t^i(x).
\end{aligned}$$

□

Remark 5.10. The symetric case of Assumption 5.8 is taking $L_i \leq 0$, and

$$q(x_t, a^i, f_t^{-i*}(x), y) - q(x_t, f_t^*(x), y) \leq 0,$$

for every $a^i \in A_i(x_t), t = T-1, \ldots, 0$. In this case, we obtain a similar result as in Proposition 5.9.

Remark 5.11. It is worth noting that basically assuming for each $i \in \bar{N}$,

$$f_t^{i*}(x) \in \arg\min \left\{ \bar{R}_t^i(a^i) \right\}, \forall x \in X, \tag{12}$$

where the strategy $f^* = \{f_t^*\}_{t=0}^{T-1}$ is as in Assumption 4.2 and

$$\bar{R}_t^i(a^i) := \int_X E_y^{f^*} \sum_{n=t+1}^{T-1} L^i(x_n, a_n^{-i}) Q(\mathrm{d}y | x, a^i, f_t^{-i*}), \tag{13}$$

the game (1) – (2) with functions r^i, P as in (11) is potential. (See Proposition 5.12, below.)

Moreover, note that for $X = \mathbb{R}^n$, $A = \mathbb{R}^m$, if the functions $\bar{R}_t^i$ are twice differentiable, then the relations

$$\begin{aligned} \frac{\partial \bar{R}_t^i}{\partial a^i}(f_t^{i*}(x)) &= 0, \text{ and} \\ \frac{\partial^2 \bar{R}_t^i}{\partial a^{2i}}(f_t^{i*}(x)) &> 0 \end{aligned} \tag{14}$$

are equivalent to (12). (Note that using convex analysis under a Gateaux or Frechet derivative criteria, the condition (12) could be also characterized, see [2]). This provides another way to obtain closed-loop potential games, as is stated in the following result.

Proposition 5.12. Under condition (12) or (14), the dynamic game (1) – (2) with functions r^i, P as in (11) is a closed-loop potential dynamic game.

P r o o f. It follows using similar arguments given in proof of Proposition 5.9. In other words, the equation $C_t^i(x, f^{i*}, f^{-i*}) = J_t^i(x)$ holds, for each $i \in \bar{N}$, $t = 0, \ldots, T-1$, $x \in X$. □

A version of (14) is considered in [18] where potential stochastic games under differentiability conditions are studied by means of the Euler equation.

See Example 5.14, below, where condition (12) is satisfied.

5.1. Another family of potential games

Observe that the dynamic game, as in (1) – (2), with functions r^i defined by

$$r^i(x, a) := p(x, a) + l^i(x, a^i), \tag{15}$$

with $p : X \times A \to \mathbb{R}$, $l^i : X \times A_i \to \mathbb{R}$ known measurable functions and

$$a = (a^1, \dots, a^{i-1}, a^i, a^{i+1}, \dots, a^N),$$

can be studied as the dynamic game (1) − (11). Indeed, take

$$\begin{aligned} P(x, a) &:= p(x, a) + \sum_{j=1}^{N} l^j(x, a^j), \\ L^i(x, a^{-i}) &:= -\sum_{j \neq i}^{N} l^j(x, a^j). \end{aligned}$$

Thus, the dynamic games (1) − (2) with cost-per-stage functions as in (15) is closed-loop potential under suitable assumptions.

In [10] is studied a deterministic version of the dynamic games (1) − (2) with cost-per-stage functions as in (15) but under an infinite horizon. In our case, we analyze the stochastic version in Section 6.

5.2. Difference equation potential stochastic games

Now, we consider a dynamic game (1) − (2) with functions r^i, P as in (11) or (15) where the state's process $\{x_t\}$ evolves according to a difference equation of the form

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t \geq 0. \tag{16}$$

The function $F : X \times A \times S \to X$ is known, x_0 is a given initial state and $\{\xi_t\}$ is a sequence of independent and identically distributed random variables taking values in a Borel space S, with a common distribution denoted by μ.

In order to illustrate this kind of games as potential games, we present the following examples.

Example 5.13. If we assume that the cost-per-stage (11) satisfies Assumption 5.3, then note that $L^i_{a_t^{-i}} := L^i(x, a_t^{-i})$ is constant in x. Thus, by Proposition 5.4, the game defined by (11) and (16) is potential. In other words, it can be solved via the OCP defined by the expected total cost functional,

$$E_x^{\pi} \sum_{t=0}^{T-1} P(x_t, a_t) + c_T(x_T),$$

subject to (16).

Similarly, if we consider the cost-per-stage (15), we require that

$$L^i(x, a^{-i}) := -\sum_{j \neq i}^{N} l^j(x, a^j)$$

satisfies Assumption 5.3 to obtain that the game defined by (15) and (16) is potential. Therefore, the OCP to solve is defined by the cost-per-stage

$$P(x,a) := p(x,a) + \sum_{j=1}^{N} l^j(x,a^j)$$

and the difference equiation (16).

Example 5.14. Consider $X = \mathbb{R}$, $A = \mathbb{R}^N$, and the functions $L^i(x,a^{-i}) := e^{-\frac{\|a^{-i}\|}{|x|}}$ and $L^i(0,a^{-i}) := 0$. We define the payoff function for the player $i \in \bar{N}$ by

$$J^i(\pi,x) := E_x^\pi \left[\sum_{t=0}^{T-1} (q \cdot x_t^2 + a_t' \cdot r \cdot a_t + L^i(x,a^{-i}) + qx_T^2 \right],$$

subject to

$$x_{t+1} = \gamma x_t + \beta a_t + \xi_t$$

where $\xi_t \in \mathbb{R}$ are independent e identical distributed random variables. Let β, and r be matrices with appropiate dimensions. Moreover, assume that r is symmetric, $\gamma \in \mathbb{R}$, $q > 0$ and r is positive definite (positive definite means $a'ra > 0$ for all $a \in \mathbb{R}^N$.) Consider the OCP defined by $P(x,a) := q \cdot x_t^2 + a_t' \cdot r \cdot a_t$. Then, by Proposition 5.12 the deterministic Markov strategy $f^* := \{f_0^*, \ldots, f_{T-1}^*\}$ with

$$f_t^*(x) := -(r + \beta' k_{t+1}\beta)^{-1}\beta' k_{t+1}\gamma x,$$

taking $k_T := q$ and for $t = T-1, \ldots, 0$,

$$k_t = \gamma[k_{t+1} - k_{t+1}\beta(r + \beta' k_{t+1}\beta)^{-1}\beta' k_{t+1}]\gamma + q,$$

is a deterministic Markov Nash equilibrium for the game. (See Section 3.5 in [7] to verify that f^* described as above is an optimal solution for the associated OCP.)

Note that the condition (12) holds because the composition

$$L^i(y, f_t^{-i*}(y)) = e^{-\|[(r+\beta' k_{t+1}\beta)^{-1}\beta' k_{t+1}\gamma]^{-i}\|}$$

is constant respect to y. Thus, for each $x \in X$,

$$f_t^{i*}(x) \in \arg\min \left\{ \bar{R}_t^i(a^i) = \sum_{n=t+1}^{T-1} e^{-\|[(r+\beta' k_{n+1}\beta)^{-1}\beta' k_{n+1}\gamma]^{-i}\|} \right\}.$$

Note that $[(r + \beta' k_{n+1}\beta)^{-1}\beta' k_{n+1}\gamma]^{-i}$ is a vector in $\mathbb{R}^{N-1}$.

Example 5.15. A particular case of difference equation games is when the players do not influence the behavior of the state's process, but the state of the system does affect their costs. In this scenario the dynamics of system (16) takes the form

$$x_{t+1} = f(x_t, \xi_t), \quad t \geq 0. \tag{17}$$

This situation occurs, for instance, when the state of the game is imposed by nature, i. e., weather, catastrophes, etc., which can not be influenced by the actions of the players. However, clearly, this kind of nature processes together with the decisions of players strongly affects the corresponding cost functions. If the game has cost functions of the form (15), then by Proposition 5.2, one can identify deterministic Markov Nash equilibria by solving the OCP given by the expected total cost functional

$$\sum_{t=0}^{T-1}[p(x_t, a_t) + \sum_{j=1}^{N} l^j(x_t, a_t^j)] + c_T(x_T),$$

subject to (17). Similarly, considering the game defined by (11) and (17) we obtain that this is a closed-loop potential stochastic game.

6. INFINITE HORIZON

In this section, we study the infinite horizon case for the dynamic stochastic game (1)–(2) with functions r^i, P as in (11) or (15). We consider the $\alpha-$discounted expected payoff function defined as

$$V_i(x, \pi) := E_x^\pi \sum_{t=0}^{\infty} \alpha^t [P(x_t, a_t) + L^i(x_t, a_t^{-i})], i \in \bar{N}, \tag{18}$$

where $\alpha \in (0, 1)$ is the so-called discount factor.

Similarly, one can take the components in (5) to define an OCP with the performance criterion defined by

$$V(x, \pi) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t P(x_t, a_t) \right], \tag{19}$$

and the value function

$$V^*(x) := \inf_{\pi \in \Pi} V(x, \pi), \quad x \in X.$$

Extending the ideas from previous sections, we impose for an infinite horizon the next assumptions.

Assumption 6.1. (See section 8.3 in [8] for existence conditions.) There exists a selector $f_0^* \in \mathbb{F}$ such that the stationary deterministic Markov strategy $f^* := (f_0^*, f_0^*, f_0^*, \ldots)$ is optimal, that is, $V(x, f^*)$ satisfies

$$\begin{aligned} V(x, f^*) &= P(x, f_0^*) + \alpha \int_X V(y, f^*) Q(\mathrm{d}y|x, f_0^*) \\ &= \min_{\Pi_{i=1}^N A_i(x)} \left[P(x, a) + \alpha \int_X V(y, f^*) Q(\mathrm{d}y|x, a) \right]. \end{aligned}$$

Assumption 6.2. For f^* as in Assumption 6.1,

$$f_0^{*i} \in argmin\{a^i \mapsto \int_X E_y^{f^*} \sum_{t=0}^{\infty} \alpha^t L^i(x_t, a_t^{-i}) Q(\mathrm{d}y|x, a^i, f_0^{-i*})\}.$$

Clearly, Assumption 6.2 is trivially satisfied when Q is an action independent transition law or when the functions L^i are constants.

Assumption 6.3. Consider f^* as in Assumption 6.1. For each $i \in \bar{N}$, and $\pi^i \in \Pi_i$,

$$\lim_{n \longrightarrow \infty} \alpha^n E_x^{\pi^i, f^{-i*}} V_i(x_n, f^*) = 0.$$

Observe that if V_i is bounded, Assumption 6.3 trivially holds.

Theorem 6.4. Under Assumptions 6.1, 6.2 and 6.3 the dynamic game (1)–(2) with functions r^i, P as in (11) or (15) is a closed-loop potential dynamic game.

Proof. Naturally, $V_i(x, f^*)$ satisfies

$$\begin{aligned} V_i(x, f^*) &= P(x, f_0^*) + L^i(x, f_0^{-i*}) + \alpha \int_X V_i(y, f^*) Q(\mathrm{d}y|x, f_0^*) \\ &= \min_{a^i \in A_i(x)} \left[P(x, a^i, f^{-i*}) + L^i(x, f^{-i*}) + \alpha \int_X V_i(y, f^*) Q(\mathrm{d}y|x, a^i, f_0^{-i*}) \right]. \end{aligned}$$

Assumption 6.3 states that f^{i*} is optimal for the corresponding OCP defined when the multi-strategy f^{-i*} is fixed. Thus, f^* is a stationary deterministic Markov Nash equilibrium for the dynamic game. □

Example 6.5. Consider a two-player game. Let $X := \{.8, .5, .2\}$, $A_i := \{0, 1\}, i = 1, 2,$ and the payoff functions described as follows,

$$\begin{aligned} r^1(x, a, b) &= \frac{x}{2}(\alpha_1 a - \alpha_2 b), \\ r^2(x, a, b) &= \frac{x}{2}(\beta_1 b - \beta_2 a). \end{aligned}$$

Assume that the constants $\alpha_i, \beta_i > 0$.

For each time t, the transition law is an arbitrary transition probability which is independent on actions, of the form $Q(x_{t+1} = y | x_t = x) := P_t^{xy} \geq 0$ such that

$$\sum_{y \in X} P_t^{xy} = 1,$$

for any $x \in X$. Let x_0 be an initial state which is known.

According to our results, the problem is reduced to maximize the function $P(x, a, b) := \frac{x}{2}(\alpha_1 a + \beta_1 b)$ for each state $x \in X$, which is attained with $f_0^{1*}(x) = 1$ and $f_0^{2*}(x) = 1$. Therefore, (f_0^{1*}, f_0^{2*}) is a deterministic Markov Nash equilibrium for the game.

7. CONCLUDING REMARKS

In this paper, we have studied a class of discrete-time stochastic games defined on Borel spaces with possibly unbounded payoffs. By means of a potential game approach together with dynamic programming arguments, it was possible to ensure the existence of deterministic Markov Nash equilibria. The importance of this results lies in the fact that in general such equilibria are formed by randomized strategies. Furthermore, our results cover, in particular, games with action independent trasition law and the so-called team games, as well as games evolving according to a stochastic difference equation.

The study of difference equation games under our approach opens up new research problems that are important from the theoretical and application points of view. Indeed, in context of works [9, 12, 15] and references there in, we can assume that the state's process $\{x_t\}$ evolves according to equation (16) or (17) where $\{\xi_t\}$ is a sequence of independent and identically distributed random variables with *unknown* distribution μ. Assuming observavility of $\{\xi_t\}$, our conjeture is that we can implement a suitable statistical estimation process for μ which combined with control schemes, defined by a potential approach, leads to obtain an approximation procedure of Markov Nash equilibria defined by a sequence of estimated-deterministic equilibria. This problem is part of a future work.

ACKNOWLEDGEMENT

This research was supported by Ciencia Frontera grant 87787 and partially supported by Consejo Nacional de Ciencia y Tecnología (CONACyT) Scholarship postdoctoral.

(Received November 12, 2021)

REFERENCES

[1] D. Dragone, L. Lambertini, G. Leitmann, and A. Palestini: Hamiltonian potential functions for differential games. Automatica *62* (2015), 134–138. DOI:10.1016/j.automatica.2015.09.036

[2] W. H. Fleming and R. W. Rishel: Deterministic and stochastic optimal control. Springer Science and Business Media *1* (2012).

[3] A. Fonseca-Morales and O. Hernández-Lerma: Potential differential games. Dyn. Games Appl. *8* (2018), 254–279. DOI:10.1007/s13235-017-0218-6

[4] A. Fonseca-Morales and O. Hernández-Lerma: Stochastic differential games: the potential approach. Stochastics *92*, (2020), 1125–1138. DOI:10.1080/17442508.2019.1691208

[5] D. González-Sánchez and O. Hernández-Lerma: Discrete-time Stochastic Control and Dynamic Potential Games: The Euler-equation Approach. Springer, New York 2013.

[6] R. Gopalakrishnan, J.Ŕ. Marden, and A. Wierman: Potential games are necessary to ensure pure Nash equilibria in cost sharing games. Math. Oper. Res. *39* (2014), 1252–1296. DOI:10.1287/moor.2014.0651

[7] O. Hernández-Lerma and J. B. Lasserre: Discrete-time Markov Control Processes: Basic Optimality Criteria. Springer-Verlag, New York 1996.

[8] O. Hernández-Lerma and J. B. Lasserre: Further Topics on Discrete-Time Markov Control Processes. Springer-Verlag, New York 1999.

[9] F. Luque-Vázquez and J. A. Minjárez-Sosa: Empirical approximation in Markov games under unbounded payoff: discounted and average criteria. Kybernetika *53* (2017), 4, 694–716. DOI:10.14736/kyb-2017-4-0694

[10] V. V. Mazalov, A. N. Rettieva, and K. E. Avrachenkov: Linear-quadratic discrete-time dynamic potential games. Autom. Remote Control *78* (2017), 1537–1544. DOI:10.1134/S0005117917080136

[11] D. Mguni: Stochastic potential games. arXiv:2005.13527v1

[12] J. A. Minjárez-Sosa: Zero-Sum Discrete-Time Markov Games with Unknown Disturbance Distribution: Discounted and Average Criteria. Springer Nature, Cham 2020. DOI:10.1007/978-3-030-35720-7

[13] D. Monderer and L. S. Shapley: Potential games. Game Econ. Behav. *14* (1996), 124–143. DOI:10.1006/game.1996.0044

[14] J. A. M. Potters, T. E. S. Raghavan, and S. H. Tijs: Pure equilibrium strategies for stochastic games via potential functions. Adv. Dyn. Games Appl. Birkhauser, Boston 2009, pp. 433–444. DOI:10.1007/978-0-8176-4834-3_22

[15] A. D. Robles-Aguilar, D. González-Sánchez, and J. A. Minjárez-Sosa: Estimation of equilibria in an advertising game with unknown distribution of the response to advertising efforts. In: Modern Trends in Controlled Stochastic Processes, Theory and Applications, V.III, (A. Piunovskiy and Y. Zhang Eds.), Springer Nature, Cham 2021. pp. 148–165. DOI:10.1007/978-3-030-76928-4_8

[16] R. W. Rosenthal: A class of games possessing pure-strategy Nash equilibria. Int. J. Game Theory *2* (1973), 65–67. DOI:10.1007/BF01737559

[17] E. M. Slade: What does an oligopoly maximize? J. Ind. Econ. *42* (1994), 45–61. DOI:10.1177/216507999404200110

[18] S. V. Macua, S. Zazo, and J. Zazo: Learning parametric closed-Loop policies for Markov potential games. arXiv:1802.00899

[19] S. Zazo, J. Zazo, and M. Sánchez-Fernández: A control theoretic approach to solve a constrained uplink power dynamic game. In: IEEE, 22nd European Signal Processing Conference on (EUSIPCO) 2014, pp. 401–405.

[20] S. Zazo, S. Valcarcel, M. Sánchez-Fernández, and J. Zazo: Dynamic potential games with constraints: fundamentals and applications in communications. IEEE Trans. Signal Proc. *64* (2016), 3806–3821. DOI:10.1109/TSP.2016.2551693

Alejandra Fonseca-Morales, Universidad de Sonora, UNISON, Hermosillo Son, 83000. México.

e-mail: alejandra.fonseca@unison.mx