KYBERNETIKA — VOLUME 57 (2021), NUMBER 3, PAGES 474–492

MARKOV STOPPING GAMES WITH AN ABSORBING STATE AND TOTAL REWARD CRITERION

Rolando Cavazos-Cadena, Luis Rodríguez-Gutiérrez
and Dulce María Sánchez-Guillermo

This work is concerned with discrete-time zero-sum games with Markov transitions on a denumerable space. At each decision time player II can stop the system paying a terminal reward to player I, or can let the system to continue its evolution. If the system is not halted, player I selects an action which affects the transitions and receives a running reward from player II. Assuming the existence of an absorbing state which is accessible from any other state, the performance of a pair of decision strategies is measured by the total expected reward criterion. In this context it is shown that the value function of the game is characterized by an equilibrium equation, and the existence of a Nash equilibrium is established.

Keywords: non-expansive operator, monotonicity property, fixed point, equilibrium equation, hitting time, bounded rewards

Classification: 91A10, 91A15

1. INTRODUCTION

This note concerns with zero-sum games with Markovian transitions on a denumerable state space and discrete-time parameter. Two players drive the system by applying actions: at each decision epoch Player II always has two options, namely, to stop the game paying a terminal reward to player I, or else, to let the system to continue its evolution, and in this case player I chooses and applies an action affecting the system transition and entitling him to receive a running reward from player II. The performance of a pair of strategies is measured by the total reward criterion and, besides standard continuity-compactness conditions, the framework of the paper is determined by the following condition, under which the total reward criterion generalizes the discounted index: There is an absorbing state which is accessible from any other state and at which the running and terminal rewards are null. In this framework, *the main objectives* of the paper can be stated as follows:

- to characterize the value function of the game via an equilibrium equation, and
- to determine a Nash equilibrium.

DOI: 10.14736/kyb-2021-3-0474

Markov stopping games with discounted criterion were studied in [5], where an application to mathematical finance was discussed; recently, in [10] the total reward criterion for *finite* models with an absorbing state was studied, and the conclusions were illustrated with two examples. As in these papers, *the approach* of this work is based on the existence of a fixed point for a non-expansive operator, a result that, via basic ideas about stopping times, dynamic programming and Markov chains, allows to derive the existence of a Nash equilibrium.

The theory of Markov games van be traced back to [15]and [23], and recent advances and applications can be found, for instance, in [1, 2, 6, 9] or [3]. On the other hand, the idea of stopping time is of great relevance in stochastic analysis, and a deep account of the theory can be found in [16] and [12]; applications to mathematical finance can be seen in [4] and [11]. On the other hand, the theory on controlled Markov processes used in this note is well established and is exposed, for instance, in [7, 8, 14], whereas applications can be found in [13, 20, 21, 22], and [17, 18, 19].

The organization of the subsequent material is as follows: The technical presentation starts at Section 2, where a Markov stopping game is formally defined, the total reward criterion is formulated, and the idea of Nash equilibrium is discussed. Then, in Section 3 a non-expansive operator is introduced and its main property is stated in Theorem 3.1, namely, that the operator admits a unique fixed point. Such a result is used to define strategies for players I and II which, as stated in Theorem 3.2, constitute a Nash equilibrium. Next, in Section 4 the basic technical tools that will be used to prove the main results are established in Lemmas 4.1–4.4. Finally, Theorems 3.1 and 3.2 are proved in Sections 5 and 6, respectively.

Notation. Given a nonempty set $\mathbb{K}$, the Banach space $\mathcal{C}(\mathbb{K})$ consist of all continuous functions $R : \mathbb{K} \to \mathbb{R}$ whose supremum norm $\|R\|$ is finite, where $\|R\| := \sup_{k\in\mathbb{K}} |R(k)|$. $\mathbb{N}$ stands for the set of nonnegative integers and the indicator function of an event A is denoted by $I[A]$. Finally, even without explicit mention, all relations involving conditional expectations are valid with probability 1 with respect to the underlying probability measure.

2. THE MODEL

This work is concerned with a dynamic model $\mathcal{G} = (S, A, \{A(x)\}_{x\in S}, R, G, P)$ which is referred to as a Markov stopping game, and whose elements have the following meaning: The (nonempty and) denumerable set S is the state space and is endowed with the discrete topology, the metric space A is the action set and, for each $x \in S$, $A(x) \subset A$ is the nonempty class of admissible actions at x for player I. On the other hand, $R \in \mathcal{C}(\mathbb{K})$ is the running reward function, where the class of admissible pairs is defined by $\mathbb{K} := \{(x, a) \,|\, a \in A(x), x \in S\}$, and $G \in \mathcal{C}(S)$ is the terminal reward; finally, $P = [p_{x,y}(a)]$ is the controlled transition law on S given $\mathbb{K}$, so that $p_{x,y}(a) \geq 0$ and $\sum_{y\in S} p_{x,y}(a) = 1$ for each $(x, a) \in \mathbb{K}$. Model $\mathcal{G}$ is interpreted as follows: At each decision epoch $t \in \mathbb{N}$, players I and II observe the state of the system, say $X_t = x \in S$, and player II must select one of two actions: To *stop* the system paying a terminal reward $G(x)$ to player I, or let the system *to continue* its evolution. In this latter case, using the record of states up to time t and actions previous to t, player I selects and applies an action (control)

$A_t = a \in A(x)$, and such an intervention has two consequences: player I gets a reward $R(x, a)$ from player II and, regardless of the previous states and actions, the system moves to $X_{t+1} = y \in S$ with probability $p_{x\,y}(a)$; this is the Markov property of the decision process. The following condition is enforced in the sequel.

Assumption 2.1. (i) For each $x \in S$, $A(x)$ is a compact subset of A.

(ii) For every $x, y \in S$, the mappings $a \mapsto R(x, a)$ and $a \mapsto p_{x,y}(a)$ are continuous in $a \in A(x)$.

(iii) For each $x \in S$ and $a \in A(x)$, $G(x) \geq 0$ and $R(x, a) \geq 0$.

Decision Strategies. For each $t = 0, 1, 2, \ldots$, the space $\mathbb{H}_t$ of possible histories up to time t is defined by $\mathbb{H}_0 := S$ and $\mathbb{H}_t := \mathbb{K}^t \times S$ when $t > 0$, whereas $\mathbf{h}_t = (x_0, a_0, \ldots, x_i, a_i, \ldots, x_t)$ stands for a generic element of $\mathbb{H}_t$, where $a_i \in A(x_i)$. A policy $\pi = \{\pi_t\}$ is a special sequence of stochastic kernels: For each $t \in \mathbb{N}$ and $\mathbf{h}_t \in \mathbb{H}_t$, $\pi_t(\cdot|\mathbf{h}_t)$ is a probability measure on A concentrated on $A(x_t)$, and for each Borel subset $B \subset A$, the mapping $\mathbf{h}_t \mapsto \pi_t(B|\mathbf{h}_t)$, $\mathbf{h}_t \in \mathbb{H}_t$, is Borel measurable. The class of all policies constitutes the family of *admissible strategies for player I* and is denoted by $\mathcal{P}$. When player I drives the system using π, the control A_t applied at time t belongs to $B \subset A$ with probability $\pi_t(B|\mathbf{h}_t)$, where $\mathbf{h}_t$ is the observed history of the process up to time t. Given $\pi \in \mathcal{P}$ and the initial state $X_0 = x$, a unique probability measure P_x^π is uniquely determined on the Borel σ-field of the space $\mathbb{H} := \prod_{t=0}^{\infty} \mathbb{K}$ of all possible realizations of the state-action process $\{(X_t, A_t)\}$ [7, 14]; the corresponding expectation operator is denoted by E_x^π. Next, define $\mathbb{F} := \prod_{x \in S} A(x)$ and notice that $\mathbb{F}$ is a compact metric space, which consists of all functions $f : S \to A$ such that $f(x) \in A(x)$ for each $x \in S$. A policy π is *stationary* if there exists $f \in \mathbb{F}$ such that the probability measure $\pi_t(\cdot|\mathbf{h}_t)$ is always concentrated at $f(x_t)$, and in this case π and f are naturally identified; with this convention, $\mathbb{F} \subset \mathcal{P}$. On the other hand, setting

$$\mathcal{F}_t := \sigma(X_0, A_0, \ldots, X_{t-1}, A_{t-1}, X_t), \tag{1}$$

the space $\mathcal{T}$ of *strategies for player II* consists of all stopping times $\tau : \mathbb{H} \to \mathbb{N}$ with respect to the filtration $\{\mathcal{F}_t\}$, that is, for each nonnegative integer t, the event $[\tau = t]$ belongs to $\mathcal{F}_t$.

Performance Criterion. Given the initial state $X_0 = x \in S$, the total expected reward of player I associated with the pair $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$ is given by

$$V(x; \pi, \tau) := E_x^\pi \left[\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) \right) I[\tau < \infty] \right] + E_x^\pi \left[I[\tau = \infty] \sum_{t=0}^{\infty} R(X_t, A_t) \right], \tag{2}$$

where the convention

$$\sum_{t=0}^{-1} R(X_t, A_t) = 0 \tag{3}$$

is enforced. When player II employs the strategy τ, the best expected total reward of player I is $\sup_{\pi\in\mathcal{P}} V(x;\pi,\tau)$, and the (*upper-*)*value* function of the game is given by

$$V^*(x) := \inf_{\tau\in\mathcal{T}}\left[\sup_{\pi\in\mathcal{P}} V(x;\pi,\tau)\right], \quad x \in S. \tag{4}$$

Interchanging the order in which the supremum and the infimum are taken, the following lower-value function of the game is obtained:

$$V_*(x) := \sup_{\pi\in\mathcal{P}}\left[\inf_{\tau\in\mathcal{T}} V(x;\pi,\tau)\right], \quad x \in S; \tag{5}$$

since $\sup_{\pi\in\mathcal{P}} V(x;\pi,\tau) \geq V(x;\pi,\tau) \geq \inf_{\tau\in\mathcal{T}} V(x;\pi,\tau)$, these definitions lead to

$$V^*(\cdot) \geq V_*(\cdot). \tag{6}$$

Equilibrium Strategies. As already mentioned, this paper is focused on the existence of a Nash equilibrium, an idea that is introduced below.

Definition 2.2. A pair $(\pi^*,\tau^*) \in \mathcal{P}\times\mathcal{T}$ is a Nash equilibrium if, for every state $x \in S$,

$$V(x;\pi,\tau^*) \leq V(x;\pi^*,\tau^*) \leq V(x;\pi^*,\tau), \quad \pi\in\mathcal{P}, \quad \tau\in\mathcal{T}. \tag{7}$$

Suppose that the strategies π^* and τ^* actually used by the players form a Nash equilibrium. In this case the first inequality in the above display shows that, if player II keeps on using strategy τ^*, then player I does not have any incentive to switch to other policy. Similarly, the second inequality in (7) implies that, if player I keeps on using π^*, then there is not any incentive for player II to change his strategy. Also, note that (7) implies that

$$V^*(\cdot) \leq \sup_{\pi} V(\cdot;\pi,\tau^*) \leq V(\cdot;\pi^*,\tau^*) \leq \inf_{\tau} V(x;\pi^*,\tau) \leq V_*(\cdot),$$

where the left- and right-most inequalities are due to (4) and (5), respectively, so that via (6) it follows that the upper and lower value functions coincide.

The existence of a Nash equilibrium was established in [5] for Markov stopping games with the discounted criterion. As it was pointed out in [10], the discounted index is a particular case of the total reward criterion applied to models with an absorbing state z satisfying two properties: (i) The running and terminal reward are null at z, and (ii) Under any stationary policy, state z is accessible from any initial state. These conditions are formally stated as follows.

Assumption 2.3. There exists a state $z \in S$ for which conditions (i)–(iii) below hold.
(i) For every $x \in S$,

$$P_x^f[\tau_z < \infty] = 1, \tag{8}$$

where

$$\tau_z := \min\{n \,|\, X_n = z\}. \tag{9}$$

(ii) $G(z) = 0 = R(z,\cdot)$ and $p_{z,z}(a) = 1, \quad a \in A(z)$.

Note that (9) yields that

$$P_z^\pi[\tau_z = 0] = 1, \quad \pi\in\mathcal{P}. \tag{10}$$

3. MAIN THEOREM

In this section the main result on the existence of a Nash equilibrium is stated. First, a subset of $\mathcal{C}(S)$ and an operator on that set are introduced.

Definition 3.1. (i) The space $[\![0, G]\!] \subset \mathcal{C}(S)$ is defined by

$$[\![0, G]\!] := \{h \in \mathcal{C}(S) \,|\, 0 \leq h(x) \leq G(x)\}. \tag{11}$$

(ii) The operator $T : [\![0, G]\!] \to [\![0, G]\!]$ is determined as follows: For every $W \in [\![0, G]\!]$ and $x \in S$,

$$T[W](x) := \min\left\{G(x), \sup_{a\in A(x)}\left[R(x, a) + \sum_{y\in S} p_{x,y}(a)W(y)\right]\right\}. \tag{12}$$

Using that R and G are nonnegative, it is not difficult to verify that T transforms $[\![0, G]\!]$ into itself and that

$$T[W](z) = W(z) = 0, \quad W \in [\![0, G]\!]. \tag{13}$$

Note that T is monotone, i. e., for $W, W_1 \in [\![0, G]\!]$,

$$W \leq W_1 \implies T[W] \leq T[W_1]. \tag{14}$$

Theorem 3.2. Under Assumptions 2.1 and 2.3 the operator T has a unique fixed point, that is, there exists a unique function $W^* \in [\![0, G]\!]$ satisfying

$$W^* = T[W^*]. \tag{15}$$

Next, strategies for players I and II will be defined using the fixed point W^*. Notice that (15) can be explicitly written as

$$W^*(x) = \min\left\{G(x), \sup_{a\in A(x)}\left[R(x, a) + \sum_{y\in S} p_{x,y}(a)W^*(y)\right]\right\}, \quad x \in S, \tag{16}$$

and observe that, since $W^* \in [\![0, G]\!]$ and G is bounded, from Assumption 2.1 it follows that there exists a policy $f^* \in \mathbb{F}$ satisfying

$$\begin{aligned} R(x, f^*(x)) + \sum_{y\in S} p_{x,y}(f^*(x))W^*(y) \\ = \sup_{a\in A(x)}\left[R(x, a) + \sum_{y\in S} p_{x,y}(a)W^*(y)\right], \quad x \in S. \end{aligned} \tag{17}$$

Next, define the subset S^* of the state space by

$$S^* := \{x \in S \,|\, W^*(x) = G(x)\}, \tag{18}$$

and let τ^* be the hitting time of set S^*, that is,

$$\tau^* := \min\{n \in \mathbb{N} \,|\, X_n \in S^*\}, \tag{19}$$

so that τ^* is a stopping time with respect to the filtration $\{\mathcal{F}_t\}$ in (1), that is, τ^* belongs to the space $\mathcal{T}$ of admissible strategies for player II. With this notation, the main conclusion of this paper can be stated as follows.

Theorem 3.3. Under Assumptions 2.1 and 2.3, the following assertions (i) – (ii) hold.
(i) For every $x \in S$,

$$V(x; f^*, \tau^*) = W^*(x);$$

(ii) The pair $(f^*, \tau^*) \in \mathbb{F} \times \mathcal{T}$ is a Nash equilibrium.

Theorem 3.2 and 3.3 will be proved in Sections 5 and 6, respectively, after the necessary preliminaries established in Section 4. The argument used to prove Theorem 3.2 relies heavily on the the monotonicity property in (14), whereas Theorem 3.3 will be established via dynamic programming arguments. Throughout the remainder Assumptions 2.1 and 2.3 are enforced.

4. TECHNICAL TOOLS

This section contains the auxiliary results that will be used to establish Theorems 3.2 and 3.3. The starting point is the following consequence of Assumption 2.1.

Lemma 4.1. (i) Consider a family $\{S_k\}$ of *finite* subsets of S such that

$$S = \bigcup_{k=1}^{\infty} S_k, \quad S_k \subset S_{k+1}, \quad k \in \mathbb{N}, \tag{20}$$

and for each $x \in S$ and $k \in \mathbb{N}$ define

$$\delta_k(x) := \sup_{a \in A(x)} \left[1 - \sum_{y \in S_k} p_{x,y}(a)\right] = \sup_{a \in A(x)} \sum_{y \in S \setminus S_k} p_{x,y}(a). \tag{21}$$

In this case,

$$\lim_{k \to \infty} \delta_k(x) = 0, \quad x \in S.$$

(ii) If $\{W_n\} \subset \mathcal{C}(S)$ is such that

$$c := \sup_{n \in \mathbb{N}} \|W_n\| < \infty \quad \text{and} \quad \lim_{n \to \infty} W_n(x) = 0, \quad x \in S, \tag{22}$$

then

$$\sup_{a \in A(x)} \sum_{y \in S} p_{x,y}(a) |W_n(y)| \to 0 \quad \text{as} \quad n \to \infty.$$

P r o o f . (i) Observe that the conditions in (20) imply that

$$\sum_{y\in S_k} p_{x,y}(a) \nearrow \sum_{y\in S} p_{x,y}(a) = 1 \text{ as } k \nearrow 1;$$

moreover, by Assumpion 2.1, for each $k \in \mathbb{N}$ the mappings $a \mapsto \sum_{y\in S_k} p_{x,y}(a)$ is continuous on the compact space $A(x)$, so that Dini's theorem implies that the above convergence is uniform on the space $A(x)$, that is, $\sup_{a\in A(x)} \left[1 - \sum_{y\in S_k} p_{x,y}(a)\right] \to 0$.
(ii) Let $x \in S$ be fixed and notice that, for every $k \in \mathbb{N}$

$$\begin{aligned}\sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a)|W_n(y)| &\le \sup_{a\in A(x)} \sum_{y\in S_k} p_{x,y}(a)|W_n(y)| + \sup_{a\in A(x)} \sum_{y\in S\setminus S_k} p_{x,y}(a)|W_n(y)| \\ &\le \max_{y\in S_k} |W_n(y)| + c \sup_{a\in A(x)} \sum_{y\in S\setminus S_k} p_{x,y}(a) \\ &= \max_{y\in S_k} |W_n(y)| + c\delta_k(x)\end{aligned}$$

where the equality is due to (21). Recalling that the sets S_k are finite, using (22) it follows that

$$\limsup_{n\to\infty} \left| \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) W_n(y) \right| \le c\delta_k(x), \quad x \in S,$$

and then, since $k \in \mathbb{N}$ is arbitrary, the conclusion follows from part (i). □

The following result establishes the continuity of the operator T with respect to the topology of pointwise convergence.

Lemma 4.2. Suppose that the sequence $\{W_n\} \subset [\![0, G]\!]$ converges pointwise to a function $V : S \to \mathbb{R}$, that is,

$$\lim_{n\to\infty} W_n(x) = V(x), \quad x \in S. \tag{23}$$

In this case

$$V \in [\![0, G]\!] \qquad \text{and} \qquad \lim_{n\to\infty} T[W_n](x) = T[V](x), \quad x \in S.$$

P r o o f . Observe that the inclusion $V \in [\![0, G]\!]$ follows from (11) and (23). Next, let $x \in S$ be arbitrary and notice that

$$\begin{aligned}&\sup_{a\in A(x)} \left[R(x,a) + \sum_{y\in S} p_{x,y}(a) W_n(y) \right] \\ &\quad = \sup_{a\in A(x)} \left[R(x,a) + \sum_{y\in S} p_{x,y}(a) V(y) + \sum_{y\in S} p_{x,y}(a)[W_n(y) - V(y)] \right] \\ &\quad \le \sup_{a\in A(x)} \left[R(x,a) + \sum_{y\in S} p_{x,y}(a) V(y) \right] + \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a)|W_n(y) - V(y)|,\end{aligned}$$

a relation that together with (12) leads to

$$T[W_n](x) \le T[V](x) + \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a)|W_n(y) - V(y)|.$$

The inequality that is obtained by interchanging the roles of W_n and V can be established along similar lines, and it follows that

$$|T[W_n](x) - T[V](x)| \le \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a)|W_n(y) - V(y)|.$$

Taking the limit as n goes to ∞ in both sides of this inequality, an application of Lemma 4.1(ii) with $|W_n - V|$ instead of W_n leads to $|T[W_n](x) - T[V](x)| \to 0$ as $n \to \infty$. □

In the second part of the following lemma, property (8) will be extended to the class of all policies.

Lemma 4.3. For each $x \in S$, and $n \in \mathbb{N}$, define

$$M_n(x) := \sup_{\pi\in\mathcal{P}} P_x^\pi[\tau_z > n] \in [0,1]. \tag{24}$$

With this notation,

(i) For each $x \in S$,

$$M_n(x) \to 0 \text{ as } n \to \infty;$$

(ii) For every $x \in S$ and $\pi \in \mathcal{P}$,

$$P_x^\pi[\tau_z < \infty] = 1.$$

P r o o f. Since $[\tau_z > n+1] \subset [\tau_z > n]$, it follows that the inequality $P_x^\pi[\tau_z > n+1] \le P_x^\pi[\tau_z > n]$ always holds, and then (24) yields that

$$M_{n+1} \le M_n, \quad n \in \mathbb{N},$$

so that

$$M(x) := \lim_{n\to\infty} M_n(x) \in [0,1] \tag{25}$$

exists for every $x \in S$; notice that (10) yields that $M_n(z) = 0$ for every positive n, so that

$$M(z) = 0. \tag{26}$$

Now, let $x \in S$ be arbitrary but fixed. Given $n \in \mathbb{N}$, select a policy $\nu_x \in \mathcal{P}$ in such a way that

$$M_{n+1}(x) - \frac{1}{n+1} \le P_x^{\nu_x}[\tau_z > n+1]. \tag{27}$$

Next, have a glance at (9) and notice that $[\tau_z > n+1] = [\tau_z > n+1, X_1 \neq z]$ for every $n \in \mathbb{N}$. With this in mind, an application of the Markov property yields that for every $\tilde{a} \in A(x)$

$$\begin{aligned} P_x^{\nu_x}[\tau_z > n+1 | A_0 = \tilde{a}] &= \sum_{y \in S\setminus\{z\}} p_{x,y}(\tilde{a}) P_y^{\nu_{x,\tilde{a}}}[\tau_z > n] \\ &\leq \sum_{y \in S\setminus\{z\}} p_{x,y}(\tilde{a}) M_n(y) \\ &\leq \sup_{a \in A(x)} \left[\sum_{y \in S\setminus\{z\}} p_{x,y}(a) M_n(y) \right], \end{aligned}$$

where policy $\nu_{x,\tilde{a}}$ is determined by $\nu_{x,\tilde{a},t}(\cdot|\mathbf{h}_t) = \nu_{x,t+1}(\cdot|x,\tilde{a},\mathbf{h}_t)$ for every $t \in \mathbb{N}$ and $\mathbf{h}_t \in \mathbb{H}_t$, and the first inequality is due to (24). Since $\tilde{a} \in A(x)$ is arbitrary, the above display leads to

$$P_x^{\nu_x}[\tau_z > n+1] \leq \sup_{a \in A(x)} \left[\sum_{y \in S\setminus\{z\}} p_{x,y}(a) M_n(y) \right].$$

On the other hand, since $M_n(\cdot) \in [0,1]$, from Assumption 2.1 it follows that there exists an action $a_{x,n} \in A(x)$ such that

$$\sum_{y \in S\setminus\{z\}} p_{x,y}(a_{x,n}) M_n(y) = \sup_{a \in A(x)} \left[\sum_{y \in S\setminus\{z\}} p_{x,y}(a) M_n(y) \right]$$

and the three previous displays together imply that

$$\begin{aligned} M_{n+1}(x) - \frac{1}{n+1} &\leq \sum_{y \in S\setminus\{z\}} p_{x,y}(a_{x,n}) M_n(y) \\ &\leq \sum_{y \in S_k\setminus\{z\}} p_{x,y}(a_{x,n}) M_n(y) + \sum_{y \in S\setminus S_k} p_{x,y}(a_{x,n}) M_n(y), \end{aligned}$$

$$M_{n+1}(x) - \frac{1}{n+1} \leq \sum_{y \in S_k\setminus\{z\}} p_{x,y}(a_{x,n}) M_n(y) + \delta_k. \tag{28}$$

Since $\{a_{x,n}\}$ is contained in the compact (metric) space $A(x)$, there exists a subsequence $\{a_{x,n_r}\}$ such that $\lim_{r\to\infty} a_{x,n_r} =: a_x \in A(x)$, and then, replacing n be n_r in the above display an taking the limit as $r \to \infty$ in both sides of the resulting inequality, Assumption 2.1 and (25) together imply that $M(x) \leq \sum_{y \in S_k\setminus\{z\}} p_{x,y}(a_x) M(y) + \delta_k$. Since this last inequality is valid for every $k \in \mathbb{N}$, via (20) and (21) it follows that

$$M(x) \leq \sum_{y \in S\setminus\{z\}} p_{x,y}(\hat{f}(x)) M(y), \quad x \in S,$$

where the stationary policy $\hat{f}$ is given by $f(x) := a_x$ for every $x \in S$. Let $n \in \mathbb{N}$ be arbitrary and, using that $[\tau_z > n+1] = [\tau_z > n, X_{n+1} \neq z]$, combine the above display with the Markov property to obtain

$$\begin{aligned} E_x^{\hat{f}}[M(X_{n+1})I[\tau_z > n+1]|\mathcal{F}_n] &= E_x^{\hat{f}}[M(X_{n+1})I[\tau_z > n, X_{n+1} \neq z]|\mathcal{F}_n] \\ &= I[\tau_z > n]E_x^{\hat{f}}[M(X_{n+1})I[X_{n+1} \neq z]|\mathcal{F}_n] \\ &= I[\tau_z > n] \sum_{y \in S\setminus\{z\}} p_{X_n,y}(\hat{f}(X_n))M(y) \\ &\geq I[\tau_z > n]M(X_n) \end{aligned}$$

where, observing that $X_n \neq z$ on the event $[\tau_n > n]$, (28) with X_n instead of x was used in the last step. It follows that the inequality

$$E_x^{\hat{f}}[M(X_n)I[\tau_z > n]] \leq |E_x^{\hat{f}}[M(X_{n+1})I[\tau_z > n+1]]$$

is always valid, and then

$$\begin{aligned} M(x)P_x^{\hat{f}}[\tau_z > 0] = E_x^{\hat{f}}[M(X_0)I[\tau_z > 0]] \leq E_x^{\hat{f}}[M(X_n)I[\tau_z > n]] \\ \leq P_x^{\hat{f}}[\tau_z > n], \quad x \in S, \quad n \in \mathbb{N}, \end{aligned}$$

where the inclusion in (25) was used to set the second inequality. Next, using that $\lim_{n\to\infty} P_x^{\hat{f}}[\tau_z > n] = 0$, by Assumption 2.3(i), the above display yields that

$$M(x)P_x^{\hat{f}}[\tau_z > 0] = 0,$$

and then, observing that $I[\tau_z > 0] = 1$ on the event $[X_0 \neq z]$, it follows that $M(x) = 0$ for $x \in S \setminus \{z\}$, so that $M(\cdot) = 0$, by (26), establishing part (i). To conclude note that for each $x \in S, n \in \mathbb{N}$ and $\pi \in \mathcal{P}$, $P_x^\pi[\tau_z > n] \leq M_n(x)$, and then part (i) implies that $P_x^\pi[\tau_z = \infty] = \lim_{n\to\infty} P_x^\pi[\tau_z > n] \leq \lim_{n\to\infty} M_n(x) = 0$, so that $P_x^\pi[\tau_z < \infty] = 1$. □

The following lemma shows that, in the context determined by Assumption 2.3, the space of strategies of player II can be reduced to the class of *finite* stopping times, a result that will be used in the proof of Theorem 3.3.

Lemma 4.4. For every $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$,

$$V(\cdot, \pi, \tau) = V(\cdot, \pi, \tau \wedge \tau_z). \tag{29}$$

P r o o f . Let $x \in S$, and $(\pi, \tau) \in \mathcal{P}\times\mathcal{T}$ be arbitrary. Keep in mind that $P_x^\pi[\tau_z < \infty] = 1$, by Lemma 4.3, and notice that Assumption 2.3 and (9) yield that

$$\text{On } [\tau_z < \infty], \quad X_{\tau_z} = z \quad \text{and} \quad R(X_n, A_n) = G(X_n) = 0, \quad n \geq \tau_z. \tag{30}$$

Next, observe the following facts (a)–(c):

(a) On the event $[\tau = \infty] \cap [\tau_z < \infty]$ the above display yields that

$$\sum_{t=0}^{\infty} R(X_t, A_t) = \sum_{t=0}^{\tau_z - 1} R(X_t, A_t)$$
$$= \sum_{t=0}^{\tau_z - 1} R(X_t, A_t) + G(X_{\tau_z}) = \sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t) + G(X_{\tau \wedge \tau_z})$$

and then, since $P_x^\pi[\tau_z < \infty] = 1$,

$$E_x^\pi \left[I[\tau = \infty] \sum_{t=0}^{\infty} R(X_t, A_t) \right] = E_x^\pi \left[I[\tau = \infty] \sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t) + G(X_{\tau \wedge \tau_z}) \right]. \quad (31)$$

(b) On the event $[\tau_z \leq \tau < \infty]$, $\tau_z = \tau \wedge \tau_z$ and via (30) it follows that $G(X_\tau) = 0 = G(X_{\tau_z}) = G(X_{\tau \wedge \tau_z})$ as well as $\sum_{t=0}^{\tau-1} R(X_t, A_t) = \sum_{t=0}^{\tau_z - 1} R(X_t, A_t) = \sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t)$ so that

$$E_x^\pi \left[I[\tau_z \leq \tau < \infty] \left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) \right) \right]$$
$$= E_x^\pi \left[I[\tau_z \leq \tau < \infty] \left(\sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t) + G(X_{\tau \wedge \tau_z}) \right) \right]$$

On the other hand, since $\tau = \tau \wedge \tau_z$ on the event $[\tau < \infty, \tau < \tau_z]$, it follows that

$$E_x^\pi \left[I[\tau < \infty, \tau < \tau_z] \sum_{t=0}^{\tau-1} R(X_t, A_t) \right]$$
$$= E_x^\pi \left[I[\tau < \infty, \tau < \tau_z]] \sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t) + G(X_{\tau \wedge \tau_z}) \right],$$

an equality that together with the previous display leads to

$$E_x^\pi \left[I[\tau < \infty] \left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) \right) \right]$$
$$= E_x^\pi \left[I[\tau < \infty] \left(\sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t) + G(X_{\tau \wedge \tau_z}) \right) \right]$$

Combining this equality with (31) and (2) it follows that

$$V(x; \pi, \tau) = E_x^\pi \left[\sum_{t=0}^{\tau \wedge \tau_z - 1} R(X_t, A_t) + G(X_{\tau \wedge \tau_z}) \right].$$

Finally, since $P_x^\pi[\tau \wedge \tau_z = \infty] = 0$, by Lemma 4.3, the right-hand side of the above display coincides with $V(x, \pi, \tau \wedge \tau_z)$, and then $V(x; \pi, \tau) = V(x; \pi, \tau \wedge \tau_z)$. □

5. THE FIXED POINT RESULT

In this section Theorem 3.2 will be established. The existence of a fixed point will be proved combining the continuity result in Lemma 4.2 with the monotonicity property (14), whereas the uniqueness will be derived via the continuity conditions in Assumption 2.1 and the optional sampling theorem.

P r o o f of Theorem 3.2. Set $W_n := T^n[0]$ for each nonnegative integer n, so that

$$W_{n+1} = T[W_n], \quad n \in \mathbb{N}. \tag{32}$$

Since $W_0 = 0 \in [\![0, G]\!]$ and $W_1 = T[0] \in [\![0, G]\!]$ it follows that $W_1 \geq W_0$, and then an induction argument combining using the above display and (14) immediately yields that

$$0 \leq W_n \leq W_{n+1} \leq G,$$

where the extreme inequalities are due to the fact that the functions W_n belong to $[\![0, G]\!]$. Thus, for each $y \in S$, the sequence $\{W_n(y)\}$ is monotone and bounded, so that

$$\lim_{n\to\infty} W_n(y) =: \hat{W}(y)$$

exists for every $y \in S$. From this point, Lemma 4.2 yields that $\hat{W} \in [\![0, G]\!]$ and

$$\lim_{n\to\infty} T[W_n](x) = T[\hat{W}](x), \quad x \in S,$$

and then, taking the limit as n goes to ∞ in both sides of (32), the two previous displays lead to

$$\hat{W} = T[\hat{W}],$$

showing that $\hat{W}$ is a fixed point if T. To conclude, it will be proved that such a fixed point is unique. Let $V \in [\![0, G]\!]$ be an arbitrary fixed point of T, so that

$$V = T[V],$$

and observe that (12) yields that for every $x \in S$

$$
\begin{aligned}
\hat{W}(x) &= T[\hat{W}](x)\\
&= \min\left\{G(x),\ \sup_{a\in A(x)}\left[R(x,a)+\sum_{y\in S}p_{x,y}(a)\hat{W}(y)\right]\right\}\\
&= \min\left\{G(x),\ \sup_{a\in A(x)}\left[R(x,a)+\sum_{y\in S}p_{x,y}(a)V(y)+\sum_{y\in S}p_{x,y}(a)[\hat{W}(y)-V(y)]\right]\right\}\\
&\le \min\left\{G(x),\ \sup_{a\in A(x)}\left[R(x,a)+\sum_{y\in S}p_{x,y}(a)V(y)\right]\right\}\\
&\qquad\qquad + \sup_{a\in A(x)}\sum_{y\in S}p_{x,y}(a)[\hat{W}(y)-V(y)]\\
&= T[V](x)+\sup_{a\in A(x)}\sum_{y\in S}p_{x,y}(a)[\hat{W}(y)-V(y)]\\
&= V(x)+\sup_{a\in A(x)}\sum_{y\in S}p_{x,y}(a)[\hat{W}(y)-V(y)],
\end{aligned}
$$

and then

$$
\hat{W}(x)-V(x)\le \sup_{a\in A(x)}\sum_{y\in S}p_{x,y}(a)[\hat{W}(y)-V(y)],\quad x\in S.
$$

Since $\hat{W}, V\in[\![0,G]\!]$ and G is bounded, it follows that $\|\hat{W}-V\|\le\|G\|<\infty$, and then Assumption 2.1 implies that there exists $f\in\mathbb{F}$ such that

$$
\sum_{y\in S}p_{x,y}(f(x))[\hat{W}(y)-V(y)]=\sup_{a\in A(x)}\sum_{y\in S}p_{x,y}(a)[\hat{W}(y)-V(y)],\quad x\in S,
$$

so that

$$
\hat{W}(x)-V(x)\le\sum_{y\in S}p_{x,y}(f(x))[\hat{W}(y)-V(y)],\quad x\in S. \tag{33}
$$

This relation and the Markov property imply that, for every $x\in S$ and $n\in\mathbb{N}$,

$$
\hat{W}(X_n)-V(X_n)\le\sum_{y\in S}p_{X_n,y}(f(X_n))[\hat{W}(y)-V(y)]=E_x^f\left[\hat{W}(X_{n+1})-V(X_{n+1})\Big|\,\mathcal{F}_n\right]
$$

and then $\{(\hat{W}(X_n)-V(X_n),\mathcal{F}_n)\}$ is a supermartingale with respect to P_x^f. Observing that $\{\hat{W}(X_n)-V(X_n)\}$ is uniformly integrable (since $\|\hat{W}-V\|\le\|G\|<\infty$) and

$$
P_x^f[\tau_z<\infty]=1,
$$

by Assumption 2.3(i), the optional sampling theorem implies that

$$
\hat{W}(x)-V(x)=E_x^f[\hat{W}(X_0)-V(X_0)]\le E_x^f[\hat{W}(X_{\tau_z})-V(X_{\tau_z})].
$$

Thus, using that $X_{\tau_z} = z$ on the event $[\tau_z < \infty]$ and $\hat{W}(z) = V(z) = 0$, by (13), the two last displays yield that $\hat{W}(x) - V(x) \leq 0$, whereas the reverse inequality can be established interchanging the roles of $\hat{W}$ and V. It follows that $\hat{W} = V$, showing that T has a unique fixed point. □

6. EXISTENCE OF NASH EQUILIBRIA

In this section a proof of Theorem 3.3 will be presented. By convenience, the core of the argument is presented separately in two lemmas.

Lemma 6.1. For each $\tau \in \mathcal{T}$,

$$W^*(\cdot) \leq V(\cdot; f^*, \tau). \tag{34}$$

P r o o f . Notice that, by Assumption 2.3 and Lemma 4.4, replacing τ by $\tau \wedge \tau_z$, if necessary, it is sufficient to establish the conclusion under the condition that τ is a finite stopping time:

$$P_x^{f^*}[\tau < \infty] = 1, \quad x \in S. \tag{35}$$

With this in mind, the verification of (34) relies on the following claim:
For every $\tau \in \mathcal{T}$ and every positive integer n and $x \in S$,

$$W^*(x) \leq E_x^{f^*}\left[\sum_{t=0}^{n-1} R(X_t, A_t) I[\tau > t] + W^*(X_\tau) I[\tau \leq n] + W^*(X_n) I[\tau > n]\right]. \tag{36}$$

Before proving this assertion, it will be used to establish the desired conclusion under the condition (35). Notice that, using the nonnegativity of R and W^* as well as (35), via the monotone convergence theorem it follows that

$$\begin{aligned}\lim_{n\to\infty} E_x^{f^*}&\left[\sum_{t=0}^{n-1} R(X_t, A_t) I[\tau > t] + W^*(X_\tau) I[\tau \leq n]\right] \\ &= E_x^{f^*}\left[\sum_{t=0}^{\infty} R(X_t, A_t) I[\tau > t] + W^*(X_\tau)\right] = E_x^{f^*}\left[\sum_{t=0}^{\tau-1} R(X_t, A_t) + W^*(X_\tau)\right].\end{aligned}$$

On the other hand, since $\|W^*| \leq \|G\| < \infty$, via (35) it follows that

$$E_x^{f^*}[W^*(X_n) I[\tau > n]] \leq \|W^*\| P_x^{f^*}[\tau > n] \to 0 \text{ as } n \to \infty.$$

Taking the limit in the right-hand side of (36) this two last displays yield that

$$\begin{aligned}W^*(x) &\leq E_x^{f^*}\left[\sum_{t=0}^{\tau-1} R(X_t, A_t) + W^*(X_\tau)\right] \\ &\leq E_x^{f^*}\left[\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)\right] = V(x, f^*, \tau), \quad x \in S,\end{aligned}$$

where the second inequality is due to the relation $W(\cdot) \leq G(\cdot)$ (derived from (16)) and, using (35), the equality is due to (2). This establishes (34). To conclude, claim (36) will be verified. To start with, note that (16) and (17) together imply that

$$W^*(x) \leq R(x, f^*(x)) + \sum_{y \in S} p_{x,y}(f^*(x)) W^*(y), \quad x \in S. \tag{37}$$

On the other hand, using that $[\tau = 0] \in \mathcal{F}_0 = \sigma(X_0)$, it follows that for each $x \in S$,

$$[X_0 = x] \subset [\tau = 0] \text{ or } [X_0 = x] \subset [\tau = 0]^c = [\tau \geq 1].$$

Now, the instance $n = 1$ of (36) will be established considering the following two exhaustive cases:

Case 1: $[X_0 = x] \subset [\tau = 0]$.

In this context $P_x^{f^*}[\tau = 0] = 1$, and it follows that

$$\begin{aligned} R(X_0) I[\tau > 0] + W^*(X_\tau) I[\tau \leq 0] + W^*(X_1) I[\tau > 0] \\ = W^*(X_0) = W^*(x) \quad P_x^{f^*}\text{-almost surely,} \end{aligned}$$

so that the inequality in (36) holds with equality when $n = 1$

Case 2: $[X_0 = x] \subset [\tau \geq 1]$.

In this framework $P_x^{f^*}[\tau > 0] = 1$, and it follows that

$$R(X_0) I[\tau > 0] + W^*(X_\tau) I[\tau \leq 0] + W^*(X_1) I[\tau > 0] = R(X_0) + W^*(X_1) \quad P_x^{f^*}\text{-almost surely,}$$

and then

$$\begin{aligned} E_x^{f^*}\left[R(X_0) I[\tau > 0] + W^*(X_\tau) I[\tau \leq 0] + W^*(X_1) I[\tau > 0] \right] \\ = E_x^{f^*}\left[R(X_0) + W(X_1) \right] \geq W^*(x) \end{aligned}$$

where the inequality is due to (37), so that the case $n = 1$ of (36) also holds in the present context.

Proceeding by induction, suppose that (36) is valid for some positive integer n and note that $W(X_n)$ and $I[\tau > n]$ are $\mathcal{F}_n$-measurable, so that

$$\begin{aligned} E_x^{f^*}[I[\tau > n] W(X_n) | \mathcal{F}_n] &= I[\tau > n] W(X_n) \\ &\leq I[\tau > n] \left(R(X_n, f^*(X_n)) + \sum_{y \in S} p_{X_n,y}(f^*(X_n)) W^*(y) \right) \\ &= E_x^{f^*}\left[R(X_n, A_n) I[\tau > n] + W^*(X_{n+1}) I[\tau > n] | \mathcal{F}_n \right] \end{aligned}$$

where (37) with X_n instead of x was used to set the inequality and the Markov property was used in the last step. Therefore,

$$\begin{aligned} &E_x^{f^*}[I[\tau > n] W(X_n)] \\ &\leq E_x^{f^*}[R(X_n, A_n) I[\tau > n]] + E_x^{f^*}[W^*(X_{n+1}) I[\tau > n]] \\ &= E_x^{f^*}[R(X_n, A_n) I[\tau > n]] + E_x^{f^*}[W^*(X_{n+1}) I[\tau = n+1]] + E_x^{f^*}[W^*(X_{n+1}) I[\tau > n+1]]. \end{aligned}$$

Combining this relation with the induction hypothesis, it follows that (36) is valid with $n+1$ instead of n, completing the induction argument. □

Lemma 6.2. For every $x \in S$

$$V(x;\pi,\tau^*) \le W^*(x), \quad \pi \in \mathcal{P}. \tag{38}$$

P r o o f . First, suppose that $x \in S^*$. In this case (18) and (19) yield that

$$W^*(x) = G^*(x) \quad \text{and} \quad P_x^\pi[\tau^* = 0] = 1,$$

whereas via (2) and (3) it follows that $V(x;\pi,\tau^*) = G(x)$, so that (38) holds with equality. To establish (38) for $x \in S \setminus S^*$, an argument similar to the one used to prove Lemma 6.1 will be used. Consider the following claim: For every $\pi \in \mathcal{P}$ and $n \in \mathbb{N} \setminus \{0\}$

$$W^*(x) \ge E_x^\pi \left[\sum_{t=0}^{n-1} R(X_t, A_t) I[\tau^* > t]] + W^*(X_{\tau^*}) I[\tau^* \le n] \right] + E_x^\pi \left[W^*(X_n) I[\tau^* > n]] \right], \quad x \in S \setminus S^*. \tag{39}$$

To verify this assertion, observe that $W^*(x) \ne G(x)$ when $x \in S \setminus S^*$, and then (16) yields that

$$\begin{aligned} W^*(x) &= \sup_{a \in A(x)} \left[R(x,a) + \sum_{y \in S} p_{x,y}(a) W^*(y) \right] \\ &\ge R(x,a) + \sum_{y \in S} p_{x,y}(a) W^*(y), \quad x \in S \setminus S^*, \quad a \in A(x); \end{aligned} \tag{40}$$

using that $P_x^\pi[\tau^* > 0] = 1$ when $x \notin S^*$, the above inequality implies that, for every $\pi \in \mathcal{P}$ and $x \in S \setminus S^*$,

$$\begin{aligned} W^*(x) &\ge E_x^\pi \left[R(X_0, A_0) + W^*(X_1) \right] \\ &= E_x^\pi \left[R(X_0, A_0) I[\tau^* > 0] + W^*(X_1) I[\tau^* = 1] + W^*(X_1) I[\tau^* > 1]] \right] \\ &= E_x^\pi \left[R(X_0, A_0) I[\tau^* > 0] + W^*(X_{\tau^*}) I[\tau^* \le 1] + W^*(X_1) I[\tau^* > 1]] \right] \end{aligned}$$

establishing (39) for the case $n = 1$. Proceeding by induction, suppose that (39) is valid for a positive integer n and, observing that $W^*(X_n)$ and $I[\tau^* > n]$ are $\mathcal{F}_n$-measurable and that $X_n \in S \setminus S^*$ on the event $[\tau^* > n]$, it follows that for every policy $\pi \in \mathcal{P}$

$$\begin{aligned} E_x^\pi [W^*(X_n) I[\tau^* > n] | \mathcal{F}_n] &= I[\tau^* > n] W(X_n) \\ &\ge I[\tau^* > n] \left(R(X_n, f^*(X_n)) + \sum_{y \in S} p_{X_n, y}(f^*(X_n)) W^*(y) \right) \\ &= E_x^\pi [R(X_n, A_n) I[\tau^* > n] + W^*(X_{n+1}) I[\tau^* > n] \,|\, \mathcal{F}_n] \end{aligned}$$

where (40) with X_n instead of x was used to set the inequality, and the Markov property was employed in the last step. Consequently,

$$\begin{aligned}
&E_x^\pi[W^*(X_n)I[\tau^* > n]\\
&\quad \geq E_x^\pi[R(X_n, A_n)I[\tau^* > n] + W^*(X_{n+1})I[\tau^* > n]]\\
&\quad \geq E_x^\pi[R(X_n, A_n)I[\tau^* > n] + W^*(X_{n+1})I[\tau^* = n+1] + W^*(X_{n+1})I[\tau^* > n+1]]\\
&\quad = E_x^\pi[R(X_n, A_n)I[\tau^* > n] + W^*(X_{\tau^*})I[\tau^* = n+1] + W^*(X_{n+1})I[\tau^* > n+1]].
\end{aligned}$$

Combining this relation with the induction hypothesis it follows that (39) holds with $n+1$ instead of n, completing the induction proof. To conclude, using that $\|W^*\| \leq \|G\| < \infty$, note that

$$E_x^\pi[W^*(X_n)I[\tau^* > n]] \to 0,$$

by Lemma 4.3, whereas the nonnegativity of R and W^* together with the monotone convergence theorem imply that

$$\begin{aligned}
\lim_{n\to\infty} E_x^\pi &\left[\sum_{t=0}^{n-1} R(X_t, A_t)I[\tau^* > t] + W^*(X_{\tau^*})I[\tau^* \leq n]\right]\\
&= E_x^\pi\left[\sum_{t=0}^{\infty} R(X_t, A_t)I[\tau^* > t] + W^*(X_{\tau^*})I[\tau^* < \infty]\right]\\
&= E_x^\pi\left[\sum_{t=0}^{\tau^*-1} R(X_t, A_t)I[\tau^* > t] + W^*(X_{\tau^*})\right].
\end{aligned}$$

Taking the limit as n goes to ∞ in the right-hand side of the inequality in (39) the two last displays yield that

$$\begin{aligned}
W(x) &\geq E_x^\pi\left[\sum_{t=0}^{\tau^*-1} R(X_t, A_t)I[\tau^* > t] + W^*(X_{\tau^*})\right]\\
&= E_x^\pi\left[\sum_{t=0}^{\tau^*-1} R(X_t, A_t)I[\tau^* > t] + G^*(X_{\tau^*})\right] = V(x; \pi, \tau^*), \quad x \in S \setminus S^*,
\end{aligned}$$

where, recalling that W and G coincide on the set S^*, the first equality is due to the inclusion $X_{\tau^*} \in S^*$, and the second equality is due to (2), showing that (38) also holds for $x \in S \setminus S^*$. □

P r o o f of Theorem 3.3. By Lemmas 6.1 and 6.2

$$V(\cdot; \pi, \tau^*) \leq W^*(\cdot) \leq V(\cdot; f^*, \tau), \quad (\pi, \tau) \in \mathcal{P} \times \mathcal{T}.$$

Setting $(\pi, \tau) = (f^*, \tau^*)$ it follows that $W^*(\cdot) = V(\cdot; f^*, \tau^*)$, and then (f^*, τ^*) is a Nash equlibrium, by Definition 2.2. □

ACKNOWLEDGEMENT

The authors are grateful to the reviewers and the Associate Editor for their careful reading of the original manuscript and for helpful suggestions to improve the paper.

(Received August 16, 2020)

REFERENCES

[1] E. Altman and A. Shwartz: Constrained Markov Games: Nash Equilibria. In: Annals of Dynamic Games (V. Gaitsgory, J. Filar and K. Mizukami, eds. *6*, Birkhauser, Boston 2000, pp. 213–221. DOI:10.1007/978-1-4612-1336-9_11

[2] R. Atar and A. Budhiraja: A stochastic differential game for the inhomogeneous ∞-Laplace equation. Ann. Probab.*2* (2010), 498–531. DOI:10.1214/09-aop494

[3] N. Bäuerle and U. Rieder: Zero-sum risk-sensitive stochastic games. Stoch. Proc. Appl. *127* (2017), 622–642. DOI:10.1016/j.spa.2016.06.020

[4] T. Bielecki, D. Hernández–Hernández, and S. R. Pliska: Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management. Math. Methods Oper. Res. *50* (1999), 167–188. DOI:10.1007/s001860050094

[5] R. Cavazos-Cadena and D. Hernández-Hernández: Nash equilibria in a class of Markov stopping games. Kybernetika *48* (2012), 1027–1044.

[6] J. A. Filar and O. J. Vrieze: Competitive Markov Decision Processes. Springer, Berlin 1996.

[7] O. Hernández-Lerma: Adaptive Markov Control Processes. Springer, New York 1989.

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

[9] V. N. Kolokoltsov and O. A. Malafeyev: Understanding Game Theory. World Scientific, Singapore 2010.

[10] V. M. Martínez-Cortés: Bipersonal stochastic transient Markov games with stopping times and total reward criteria. Kybernetika *57* (2021), 1–14. DOI:10.14736/kyb-2021-1-0001

[11] G. Peskir: On the American option problem. Math. Finance *15* (2005), 169–181. DOI:0.5840/leibniz20051510

[12] G. Peskir and A. Shiryaev: Optimal Stopping and Free-Boundary Problems. Birkhauser, Boston 2010.

[13] A. B. Piunovskiy: Examples in Markov Decision Processes. Imperial College Press, London 2013.

[14] M. Puterman: Markov Decision Processes. Wiley, New York 1994.

[15] L. S. Shapley: Stochastic games. Proc. Natl. Acad. Sci. USA *39* (1953), 1095–1100.

[16] A. Shiryaev: Optimal Stopping Rules. Springer, New York 1978.

[17] K. Sladký: Ramsey growth model under uncertainty. In: Proc. 27th International Conference Mathematical Methods in Economics (H. Brozová, ed.), Kostelec nad Černými lesy 2009, pp. 296–300.

[18] K. Sladký: Risk-sensitive Ramsey growth model. In: Proc. 28th International Conference on Mathematical Methods in Economics (M. Houda and J. Friebelová, eds.) České Budějovice 2010, pp. 560–565.

[19] K. Sladký: Risk-sensitive average optimality in Markov decision processes. Kybernetika *54* (2018), 1218–1230. DOI:10.14736/kyb-2018-6-1218

[20] D. J. White: Real applications of Markov decision processes. Interfaces *15* (1985), 73–83. DOI:10.1287/inte.15.6.73

[21] D. J. White: Further real applications of Markov decision processes. Interfaces *18* (1988), 55–61. DOI:10.1287/inte.18.5.55

[22] D. J. White: A survey of applications of Markov decision processes. J. Opl. Res. Soc. *44* (1993), 1073–1096. DOI:10.1016/0042-207X(93)90304-S

[23] L. E. Zachrisson: Markov Games. In: Advances in Game Theory (M. Dresher, L. S. Shapley and A. W. Tucker, eds.), Princeton Univ. Press, Princeton N.J. 1964, pp. 211–253.

Rolando Cavazos-Cadena, Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Boulevard Antonio Narro 1923, Buenavista, Saltillo, COAH 25315. México.

e-mail: rolando.cavazos@uaaan.edu.mx

Luis Rodríguez-Gutiérrez , Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Boulevard Antonio Narro 1923, Buenavista, Saltillo, COAH 25315. México.

e-mail: lrodgut@hotmail.com

Dulce María Sánchez-Guillermo, Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Boulevard Antonio Narro 1923, Buenavista, Saltillo, COAH 25315. México.

e-mail: ddulcemar3@gmail.com