Kybernetika

Rolando Cavazos-Cadena; Daniel Hernández-Hernández

Nash Equilibria in a class of Markov stopping games

Kybernetika, Vol. 48 (2012), No. 5, 1027--1044

Persistent URL: `http://dml.cz/dmlcz/143097`

KYBERNETIKA — VOLUME 48 (2012), NUMBER 5, PAGES 1027–1044

NASH EQUILIBRIA IN A CLASS OF MARKOV STOPPING GAMES

Rolando Cavazos-Cadena and Daniel Hernández-Hernández

This work concerns a class of discrete-time, zero-sum games with two players and Markov transitions on a denumerable space. At each decision time player II can stop the system paying a terminal reward to player I and, if the system is no halted, player I selects an action to drive the system and receives a running reward from player II. Measuring the performance of a pair of decision strategies by the total expected discounted reward, under standard continuity-compactness conditions it is shown that this stopping game has a value function which is characterized by an equilibrium equation, and such a result is used to establish the existence of a Nash equilibrium. Also, the method of successive approximations is used to construct approximate Nash equilibria for the game.

Keywords: zero-sum stopping game, equality of the upper and lower value functions, contractive operator, hitting time, stationary strategy

Classification: 91A10, 91A15

1. INTRODUCTION

This note is concerned with a class of discrete time, zero-sum games evolving on a denumerable state space according to a (time invariant) Markovian transition mechanism. There are two players in the game and, at each observation time $t = 0, 1, 2, 3, \ldots$, they consider the previous history as well as the current state to select their decisions: player II can stop the game paying a terminal reward to player I, or can allow the system to continue its evolution, and in this latter case player I chooses an action influencing the system transition and entitling him to receive a running reward from player II. It is assumed that the available actions for player I form a compact metric space at each state, and that the running reward and the system transitions depend continuously on the applied action. The performance of a pair of decision strategies is measured by the total expected discounted reward of player I and, within this context, *the main conclusions* of the paper are as follows: (a) It is shown that the upper and lower value functions of this stopping game are the same, say V^*, and (b) an equilibrium equation characterizing V^* is derived. Next, (c) using such an equation the existence of a Nash equilibrium for the game is established and, finally (d) it is shown that a successive approximation scheme can be used to determine strategies for both players which form an 'approximate' Nash equilibrium in a sense to be formally specified below.

The theory of games has interesting applications in diverse areas; see, for instance, Altman and Schwartz [1], Atar and Budhiraja [2], and the recent book by Kolokoltsov and Malafeyev [5], and it should be mentioned that the topic of Markov Games was initiated in the pioneer papers by Shapley [12] and Zachrisson [17]. Also, stopping time problems have been intensively studied, and a fairly complete account of the theory can be found in Shiryaev [9] and Peskir and Shiryaev [7], whereas an application to mathematical finance was presented in Peskir [6]. An intersection between game theory and optimal stopping was presented in Dynkin [4], where games with two players were studied and each player can stop the system, and in van der Wal [13, 14].

On the other hand, the field of (risk-neutral) controlled Markov chains has a well-established theory (Puterman [8]), and good account of real successful applications can be found, for instance, in White [15, 16], whereas recent applications to economic growth problems can be found in Sladký [10, 11], where discounted and risk sensitive criteria were studied. In this note, the fundamental ideas of optimal stopping and Markov decision processes are combined to analyze the stopping game described above.

The organization of the subsequent material is as follows: In Section 2 the Markov stopping game model is formally introduced, and the classes of strategies for both players, the idea of a Nash equilibrium for the game, as well as the discounted criterion are briefly discussed. Next, in Section 3 it is shown that the value function of the game is well-defined and is characterized as the unique bounded solution of an equilibrium equation, which is used in Section 4 to establish the existence of a Nash equilibrium. Finally, in Section 5 the idea of ε-Nash equilibrium is introduced and it is shown that, for each $\varepsilon > 0$, a successive approximation method can be used to construct ε-Nash equilibria for the stopping game, and the exposition concludes with a brief discussion of an example in Section 6 before the bibliography.

Notation. Given a topological space $\mathbb{K}$, the Banach space $\mathcal{B}(\mathbb{K})$ consist of all continuous functions $R\colon \mathbb{K} \to \mathbb{R}$ whose supremum norm $\|R\|$ is finite, where $\|R\| := \sup_{k\in\mathbb{K}} |R(k)|$, whereas

$$\mathbb{N} := \{0, 1, 2, 3, \ldots\} \cup \{\infty\}.$$

On the other hand, for numbers $r_0, r_1, r_2, \ldots,$

$$\sum_{t=0}^{-1} r_t = 0,$$

and the indicator function of an event A is denoted by $I[A]$. Finally, without explicit mention, all relations involving conditional expectations are valid with probability 1 with respect to the underlying probability measure.

2. THE MODEL

Throughout the remainder $\mathcal{G} = (S, A, \{A(x)\}_{x\in S}, R, G, P)$ stands for a zero-sum stopping sequential game with two players I and II, where the state space S is a denumerable set endowed with the discrete topology, and the action set A is a metric space. For each

$x \in S$, $A(x) \subset A$ is the nonempty set of admissible actions at x for player I, whereas $\mathbb{K} := \{(x,a) \,|\, a \in A(x), x \in S\}$ is the class of admissible pairs, which is considered as a topological subspace of $S \times A$. On the other hand, $R \in \mathcal{B}(\mathbb{K})$ and $G \in \mathcal{B}(S)$ are the running and terminal reward functions, respectively, whereas $P = [p_{x\,y}(\cdot)]$ is the controlled transition law on S given $\mathbb{K}$, that is, $p_{x\,y}(a) \geq 0$ and $\sum_{y\in S} p_{x\,y}(a) = 1$ for each $(x,a) \in \mathbb{K}$. This model $\mathcal{G}$ is interpreted as follows: At each time $t = 0, 1, 2, \ldots$, players I and II observe the state of the system, say $X_t = x \in S$, and player II can decide to stop the system at the expense of paying a terminal reward $G(x)$ to player I, or else player II can decide to let the system to continue its evolution. In this latter case, player I uses the history of previously observed states and actions applied, as well as the current state $X_t = x$, to select an action (control) $A_t = a \in A(x)$ to drive the system. As a consequence, player I gets a reward $R(x,a)$ from player II and, regardless of the previous states and actions, the state of the system at time $t+1$ will be $X_{t+1} = y \in S$ with probability $p_{x\,y}(a)$; this is the Markov property of the sequential decision process. From this point onwards, the following condition is enforced even without explicit reference.

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 $a \in A(x)$.

Decision Strategies. The space $\mathbb{H}_t$ of possible histories up to time $t = 0, 1, 2, \ldots$ is defined by $\mathbb{H}_0 := S$ and $\mathbb{H}_t := \mathbb{K}^t \times S$ for $t = 1, 2, 3, \ldots$; a generic element of $\mathbb{H}_t$ is denoted by $\mathbf{h}_t = (x_0, a_0, \ldots, x_i, a_i, \ldots, x_t)$, where $a_i \in A(x_i)$. A policy $\pi = \{\pi_t\}$—or decision strategy for player I—is a special sequence of stochastic kernels: For each $t = 0, 1, 2, \ldots$ 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; when the system is driven according to 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. The class of all policies is denoted by $\mathcal{P}$. Given the policy π and the initial state $X_0 = x$, a unique probability measure P_x^π is uniquely determined in the product space

$$\mathbb{H} := \prod_{t=0}^{\infty} \mathbb{K}$$

of all possible realizations of the state-action process $\{(X_t, A_t)\}$ (Puterman [8]); 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 a function $f \in \mathbb{F}$ 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 $\mathcal{F}_t := \sigma(X_0, A_0, \ldots, X_{t-1}, A_{t-1}, X_t)$, 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 $\mathcal{F}_t$.

Discounted Value Functions and Nash Equilibria. Let $\beta \in (0,1)$ be a discount factor which will be held fixed throughout the reminder. Given an initial state $x \in S$, the total expected discounted reward of player I corresponding to the pair $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$ is given by

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

where, by convention,

$$\beta^\tau G(X_\tau) = 0 \quad \text{on the event } [\tau = \infty]. \tag{2}$$

Notice that

$$|V_\beta(x; \pi, \tau)| \leq E_x^\pi \left[\sum_{t=0}^{\infty} \beta^t |R(X_t, A_t)| + \|G\| \right] \leq \frac{\|R\|}{1-\beta} + \|G\| < \infty,$$

so that $V_\beta(x; \pi, \tau)$ is always well-defined and satisfies

$$\|V_\beta(\cdot; \pi, \tau)\| \leq \frac{\|R\|}{1-\beta} + \|G\|. \tag{3}$$

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

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

combining the two last displays it follows that $V_\beta^* \in \mathcal{B}(S)$.

Remark 2.1. The discounted *lower-value* function of the game is specified by interchanging the order in which the supremum and infimum are taken in (4):

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

Observing that the relation

$$\sup_{\pi \in \mathcal{P}} V_\beta(x; \pi, \tau) \geq V_\beta(x; \pi, \tau) \geq \inf_{\tau \in \mathcal{T}} V_\beta(x; \pi, \tau)$$

is always valid, it follows from (4) and (5) that the inequality $V_\beta^*(x) \geq V_{\beta,*}(x)$ holds for every state x. As it will be shown below, under Assumption 2.1 the value functions $V_\beta^*(\cdot)$ and $V_{\beta,*}(\cdot)$ coincide.

Definition 2.1. A pair $(\pi^*, \tau^*) \in \mathcal{P} \times \mathcal{T}$ is a Nash equilibrium if

$$V_\beta(x; \pi, \tau^*) \leq V_\beta(x; \pi^*, \tau^*), \quad x \in S, \quad \pi \in \mathcal{P},$$

and

$$V_\beta(x, \pi^*, \tau) \geq V_\beta(x, \pi^*, \tau^*), \quad x \in S, \quad \tau \in \mathcal{T}.$$

Since the reward of player I is received from player II, it follows that when the actual strategies π^* and τ^* used by the players form a Nash equlibrium, then if one player keeps on using his strategy, the other one does not have any incentive to change her/his behaviour.

The Problems. The main objectives of this note can be described as follows:

1. To establish an equilibrium equation characterizing the value function $V_\beta^*(\cdot)$;

2. To show that such an equation renders a Nash equilibrium (π^*, τ^*) for the game, and

3. To prove that, when the players follow the above strategies π^* and τ^*, the expected discounted reward obtained by player I coincides with the value function $V_\beta^*(\cdot)$, that is,

$$V_\beta^*(x) = V_\beta(x; \pi^*, \tau^*), \quad x \in S.$$

4. To implement, starting from scratch, an algorithm allowing to obtain strategies $\hat{\pi}$ and $\hat{\tau}$ which form an approximate Nash equilibrium, in a sense to be precisely formulated below.

3. EQUILIBRIUM EQUATION

In this section an equilibrium equation characterizing the value function V_β^* in (4) will be established. To begin with, define the operator $C : \mathcal{B}(S) \to \mathcal{B}(S)$ as follows: For each $W \in \mathcal{B}(S)$ and $x \in S$,

$$C[W](x) := \min\left\{ G(x), \sup_{a \in A(x)} \left[R(x,a) + \beta \sum_{y \in S} p_{x\,y}(a) W(y) \right] \right\}. \tag{6}$$

Notice now that C is monotone and β-subhomogeneous, that is, for $W, W_1 \in \mathcal{B}(S)$,

$$C[W] \le C[W_1] \quad \text{if } W \le W_1, \tag{7}$$

and

$$C[W + r] \le C[W] + \beta r, \qquad \text{if } r \in [0, \infty).$$

Observing that the relation $W_1 \le W_2 + \|W_2 - W_1\|$ is always valid for $W_1, W_2 \in \mathcal{B}(S)$, these properties immediately yield that $C[W_1] \le C[W_2] + \beta\|W_1 - W_2\|$; interchanging he roles of W_1 and W_2, it follows that

$$\|C[W_1] - C[W_2]\| \le \beta\|W_1 - W_2\|, \quad W_1, W_2 \in \mathcal{B}(S), \tag{8}$$

that is, C is a contractive operator in the Banach space $\mathcal{B}(S)$, and then it has a unique fixed point W^*,

$$C[W^*] = W^*, \quad W^* \in \mathcal{B}(S). \tag{9}$$

Moreover, such a fixed point W^* can be obtained as the uniform limit of successive compositions of C. More explicitly, for each positive integer n and $W \in \mathcal{B}(S)$, $\|W^* - C^n[W]\| \le \beta^n \|W^* - W\|$, and then

$$W^* = \lim_{n \to \infty} C^n[W], \quad W \in \mathcal{B}(S). \tag{10}$$

Theorem 3.1. Suppose that Assumption 2.1 holds, and let the (upper and lower discounted) value functions $V_\beta^*(\cdot)$ and $V_{\beta,*}(\cdot)$ be as in (4) and (5), respectively. In this case the following assertions (*i*)–(*iii*) are valid, where the operator C is as in (6), and W^* is the corresponding fixed point.

(*i*) $V_\beta^* \le C[V_\beta^*]$, and

(*ii*) $V_{\beta,*} \ge C[V_{\beta,*}]$.

Consequently,

(*iii*) $V_\beta^* = W^* = V_{\beta,*}$, and then V_β^* is the unique solution in $\mathcal{B}(S)$ of the following equilibrium equation:

$$V_\beta^*(x) = \min\left\{ G(x), \sup_{a\in A(x)} \left[R(x,a) + \beta \sum_{y\in S} p_{x\,y}(a) V_\beta^*(y)\right]\right\}, \quad x\in S. \tag{11}$$

Proof. (i) Let π be an arbitrary strategy for player I, and let $\tau \equiv 0$. In this case, $V_\beta(x;\pi,\tau) = G(x)$, by (1), so that $\sup_{\pi\in\mathcal{P}} V_\beta(x;\pi,\tau) = G(x)$, and then (4) yields that

$$V_\beta^*(x) \le G(x), \quad x\in S. \tag{12}$$

Next, let $\varepsilon > 0$ be arbitrary. Using (4), for each $y\in S$ select a stopping time $\tau_y : \mathbb{H} \to \mathbb{N}$ such that

$$\sup_{\delta\in\mathcal{P}} V_\beta(y;\delta,\tau_y) \le V_\beta^*(y) + \varepsilon, \tag{13}$$

and define the new stopping time $\tilde\tau : \mathbb{H} \to \mathbb{N}$ as follows: For $\mathbf{h} = (x_0, a_0, x_1, a_1, \ldots) \in \mathbb{H}$

$$\tilde\tau(\mathbf{h}) = 1 + \tau_{x_1}(x_1, a_1, x_2, a_2, \ldots).$$

In words, when player II stops the system according to $\tilde\tau$, the system runs at least up to time 1, and if $X_1 = y$ is observed, then the system is halted at time $1+k$, where k is the value attained by τ_y as if the process had started again at time 1. Now, given $(x,a)\in\mathbb{K}$, define the shifted strategy $\pi^{(x,a)}$ as follows: for each nonnegative integer t, $\pi_t^{(x,a)}(\mathbf{h}_t) = \pi_{t+1}(x,a,\mathbf{h}_t)$. When player I chooses actions according to $\pi^{(x,a)}$, he prefixes the observed history $\mathbf{h}_t$ with the pair (x,a) and then selects actions according to the original policy π as if the augmented history $(x,a,\mathbf{h}_t)$ had been observed. With this notation, an application of the Markov property yields that, for every $x\in S$,

$$\begin{aligned} E_x^\pi &\left[\sum_{t=0}^{\tilde\tau - 1} \beta^t R(X_t, A_t) + \beta^{\tilde\tau} G(X_{\tilde\tau}) \,\middle|\, A_0 = a, X_1 = y\right] \\ &= R(x,a) + \beta E_y^{\pi^{(x,a)}} \left[\sum_{t=0}^{\tau_y - 1} \beta^t R(X_t, A_t) + \beta^{\tau_y} G(X_{\tau_y})\right] \\ &= R(x,a) + \beta V_\beta(y; \pi^{(x,a)}, \tau_y) \\ &\le R(x,a) + \beta \sup_{\delta\in\mathcal{P}} V_\beta(y;\delta,\tau_y) \\ &\le R(x,a) + \beta[V_\beta^*(y) + \varepsilon], \end{aligned}$$

where (13) was used to set the last inequality. Taking the expectation with respect to X_1 it follows that

$$\begin{aligned} E_x^\pi \left[\sum_{t=0}^{\tilde\tau-1} \beta^t R(X_t, A_t) + \beta^{\tilde\tau} G(X_{\tilde\tau}) \middle| A_0 = a, \right] &\\ \leq \ R(x,a) + \beta \sum_{y \in S} p_{x\,y}(a) V_\beta^*(y) + \beta\varepsilon &\\ \leq \ \sup_{a \in A(x)} \left[R(x,a) + \beta \sum_{y \in S} p_{x\,y}(a) V_\beta^*(y) \right] + \beta\varepsilon & \end{aligned}$$

and then

$$\begin{aligned} V_\beta(x;\pi,\tilde\tau) &= E_x^\pi \left[\sum_{t=0}^{\tilde\tau-1} \beta^t R(X_t, A_t) + \beta^{\tilde\tau} G(X_{\tilde\tau}) \right] \\ &\leq \sup_{a \in A(x)} \left[R(x,a) + \beta \sum_{y \in S} p_{x\,y}(a) V_\beta^*(y) \right] + \beta\varepsilon. \end{aligned}$$

Since $\pi \in \mathcal{P}$ and $\varepsilon > 0$ are arbitrary, it follows that

$$V_\beta^*(x) \leq \sup_{\pi \in \mathcal{P}} V_\beta(x;\pi,\tilde\tau) \leq \sup_{a \in A(x)} \left[R(x,a) + \beta \sum_{y \in S} p_{x\,y}(a) V_\beta^*(y) \right],$$

where the first inequality stems from (4). Combining this relation with (12) and the specification of the operator C in (6), it follows that $V_\beta^* \leq C[V_\beta^*]$.

(ii) Let $\varepsilon > 0$ be arbitrary and, using (5), for each $y \in S$ select a strategy $\pi^y \in \mathcal{P}$ such that

$$\inf_{\tilde\tau \in \mathcal{T}} V_\beta(y;\pi^y,\tilde\tau) \geq V_{\beta,*}(y) - \varepsilon. \tag{14}$$

Next, observe that $V_{\beta,*}$ is a bounded function (see (3) and (5)) so that, by Assumption 2.1, there exists a stationary strategy $f \in \mathbb{F}$ such that

$$\begin{aligned} R(x,f(x)) + \beta \sum_{y \in S} p_{x\,y}(f(x)) V_{\beta,*}(y) &\\ = \sup_{a \in A(x)} \left[R(x,a) + \beta \sum_{y \in S} p_{x\,y}(a) V_{\beta,*}(y) \right], \quad x \in S. & \end{aligned} \tag{15}$$

Now, this stationary strategy f and the strategies π^y in (14) will be used to construct a new policy $\pi^f \in \mathcal{P}$ as follows: For each nonnegative integer t and $\mathbf{h}_t = (x_0, a_0, \ldots, x_t) \in \mathbb{H}_t$,

$$\begin{aligned} &\pi_0^f(\mathbf{h}_0) = f(x_0), \quad \pi_1^f(\mathbf{h}_1) = \pi_0^{x_1}(x_1) \\ &\pi_t^f(\mathbf{h}_t) = \pi_{t-1}^{x_1}(x_1, a_1, \ldots, x_t), \quad t \geq 2. \end{aligned} \tag{16}$$

In words, when player I follows the strategy π^f, actions are selected according to f at time 0 and, if $X_1 = y$ is observed, from time 1 onwards actions are chosen according to π^y as if the process had started again. To continue, let $\tau : \mathbb{H} \to \mathbb{N}$ be an arbitrary stopping time and, recalling that the event $[\tau = 0]$ belongs to the σ-field $\mathcal{F}_0 = \sigma(\mathbb{H}_0) = \sigma(X_0)$, observe that *exactly one* of the following statements is valid for each $x \in S$:

$$[X_0 = x] \subset [\tau = 0], \quad \text{or} \quad [X_0 = x] \cap [\tau = 0] = \emptyset.$$

With this in mind, for a given state $x \in S$, the discounted reward function $V_\beta(x; \pi^f, \tau)$ will be analyzed. Consider the following two exhaustive cases (a) and (b):

(a) The inclusion $[X_0 = x] \subset [\tau = 0]$ occurs: In this context, it follows that $1 = P_x^{\pi^f}[X_0 = x] = P_x^{\pi^f}[\tau = 0] = 1$, and then

$$V_\beta(x; \pi^f, \tau) = G(x).$$

(b) The events $[X_0 = x]$ and $[\tau = 0]$ are disjoint: In this case $P_x^{\pi^f}[\tau \geq 1] = 1$ and an application of the Markov property using the specification of the policy π^f in (16) yields that

$$V_\beta(x; \pi^f, \tau) = R(x, f(x)) + \beta \sum_{y \in S} p_{x\,y}(f(x)) V_\beta(y; \pi^y, \hat{\tau}),$$

where the stopping time $\hat{\tau} : \mathbb{H} \to \mathbb{N}$ is given by $\hat{\tau}(\mathbf{h}) = \tau(x, f(x), \mathbf{h}) - 1$, $\mathbf{h} \in \mathbb{H}$. Therefore,

$$\begin{aligned} V_\beta(x; \pi^f, \tau) &\geq R(x, f(x)) + \beta \sum_{y \in S} p_{x\,y}(f(x)) \inf_{\tau \in \mathcal{T}} V_\beta(y; \pi^y, \tau) \\ &\geq R(x, f(x)) + \beta \sum_{y \in S} p_{x\,y}(f(x)) V_{\beta,*}(y) - \beta\varepsilon \\ &= \sup_{a \in A(x)} \left[R(x, a) + \beta \sum_{y \in S} p_{x\,y}(a) V_{\beta,*}(y) \right] - \beta\varepsilon, \end{aligned}$$

where (14) and (15) were used in the two last steps.

Combining the conclusions obtained in the analysis of the above cases (a) and (b) with the specification of the contractive operator C in (6), it follows that for every $\tau \in \mathcal{T}$ and $x \in S$

$$V_\beta(x; \pi^f, \tau) \geq C[V_{\beta,*}](x) - \beta\varepsilon,$$

so that

$$V_{\beta,*}(x) \geq \inf_{\tau \in \mathcal{T}} V_\beta(x; \pi^f, \tau) \geq C[V_{\beta,*}](x) - \beta\varepsilon, \quad x \in S;$$

see (5) for the first inequality. Since $\varepsilon > 0$ is arbitrary, it follows that $V_{\beta,*} \geq C[V_{\beta,*}]$.

(iii) *Via* the monotonicity property of C in (7), the two previous parts yield that $V_{\beta,*} \geq C^n[V_{\beta,*}]$ and $C^n[V_\beta^*] \geq V_\beta^*$ for every positive integer n, and then (10) leads to $V_{\beta,*} \geq W^*$ and $W^* \geq V_\beta^*$; since the value functions satisfy $V_\beta^* \geq V_{\beta,*}$, it follows that

$$V_{\beta,*} = W^* = V_\beta^*.$$

From this point, the specification of the operator C in (6) yields that the equilibrium equation (11) is equivalent to (9), concluding the argument. □

4. NASH EQULIBRIA

In this section strategies f^* and τ^* for players I and II will be specified using the value function V^*_β, and it will be shown that the pair (f^*, τ^*) is a Nash equlibrium. To begin with, let $\mathcal{S}^*$ be the subset of states where the value function V^*_β and the terminal reward G coincide, that is,

$$\mathcal{S}^* := \{x \in S \,|\, G(x) = V^*_\beta(x)\}, \tag{17}$$

and let τ^* be the first arrival time to $\mathcal{S}^*$, that is,

$$\tau^*(\mathbf{h}) := \min\{t \geq 0 \,|\, x_t \in \mathcal{S}^*\}, \quad \mathbf{h} = (x_0, a_0, x_1, a_1, \ldots) \in \mathrm{I\!H}. \tag{18}$$

Now, let $f^* \in \mathrm{I\!F}$ be a the stationary strategy satisfying

$$\begin{aligned} R(x, f^*(x)) &+ \beta \sum_{y \in S} p_{x\,y}(f^*(x)) V^*_\beta(y) \\ &= \sup_{a \in A(x)} \left[R(x, a) + \beta \sum_{y \in S} p_{x\,y}(a) V^*_\beta(y) \right], \quad x \in S, \end{aligned} \tag{19}$$

whose existence is guaranteed by Assumption 2.1.

Theorem 4.1. The pair of strategies $(f^*, \tau^*) \in \mathrm{I\!F} \times \mathcal{T}$ described above is a Nash equilibrium.

The proof of this result relies in the following lemma, establishing that when players I and II use the strategies f^* and τ^*, respectively, the expected discounted reward of player I coincides with the value function V^*_β.

Lemma 4.2. For each $x \in S$, the strategies (f^*, τ^*) in (18) and (19) satisfy

$$V^*_\beta(x) = V_\beta(x; f^*, \tau^*); \tag{20}$$

see (1) and (4).

Proof. Suppose that $x \in \mathcal{S}^*$, so that

$$G(x) = V^*_\beta(x),$$

by (17). On the other hand, observing $\tau^* = 0$ on the event$[X_0 = x]$, it follows that $1 = P^{f^*}_x[X_0 = x] = P^{f^*}_x[\tau^* = 0]$, and then

$$V_\beta(x, f^*, \tau^*) = E^{f^*}_x \left[\sum_{t=0}^{\tau^*-1} \beta^t R(X_t, A_t) + \beta^{\tau^*} G(X_{\tau^*}) \right] = G(x),$$

establishing that (20) holds when x belongs to $\mathcal{S}^*$. To verify the desired conclusion when x lays outside of $\mathcal{S}^*$, observe that (17)–(19) together with the equilibrium equation (11) yield that

$$V_\beta^*(x) = R(x, f^*(x)) + \beta \sum_{y\in S} p_{x\,y}(f^*(x))V_\beta^*(y), \quad x \in S \setminus \mathcal{S}^*, \tag{21}$$

that is,

$$V_\beta^*(x) = E_x^{f^*}\left[R(X_0, A_0) + \beta I[\tau^* \geq 1]V_\beta^*(X_1)\right], \quad x \in S \setminus \mathcal{S}^*, \tag{22}$$

where it was used that the inequality $\tau^* \geq 1$ occurs $P_x^{f^*}$-almost surely if the initial state x does not belong to $\mathcal{S}^*$, by (18). It will be shown that, for each positive integer n,

$$\begin{aligned} V_\beta^*(x) \quad = \quad & E_x^{f^*}\left[\sum_{t=0}^{n-1} \beta^t R(X_t, A_t) I[\tau^* > t]\right] \\ & + E_x^{f^*}\left[\beta^{\tau^*} G(X_{\tau^*}) I[\tau^* < n]\right] \\ & + E_x^{f^*}\left[\beta^n I[\tau^* \geq n] V_\beta^*(X_n)\right], \quad x \in S \setminus \mathcal{S}^*. \end{aligned} \tag{23}$$

To establish this claim notice that, since $P_x^{f^*}[\tau^* \geq 1] = 1$ when $x \in S \setminus \mathcal{S}^*$, this last assertion reduces to (22) when $n = 1$. Proceeding by induction, assume now that (23) holds for some positive integer n, let $x \in S \setminus \mathcal{S}^*$ be arbitrary and notice that

$$\begin{aligned} & E_x^{f^*}\left[\beta^n I[\tau^* \geq n] V_\beta^*(X_n)\right] \\ & \quad = E_x^{f^*}\left[\beta^n I[\tau^* = n] V_\beta^*(X_n)\right] + E_x^{f^*}\left[\beta^n I[\tau^* \geq n+1] V_\beta^*(X_n)\right] \\ & \quad = E_x^{f^*}\left[\beta^{\tau^*} I[\tau^* = n] G(X_{\tau^*})\right] + E_x^{f^*}\left[\beta^n I[\tau^* \geq n+1] V_\beta^*(X_n)\right] \end{aligned} \tag{24}$$

where, observing that $X_{\tau^*} \in \mathcal{S}^*$ when τ^* is finite, the second equality is due to the fact that G and V_β^* coincide on $\mathcal{S}^*$. Next, notice that $[\tau^* \geq n+1] = [\tau^* \leq n]^c \in \mathcal{F}_n$, and then

$$\begin{aligned} & E_x^{f^*}\left[\beta^n I[\tau^* \geq n+1] V_\beta^*(X_n) \middle| \mathcal{F}_n\right] \\ & \quad = \beta^n I[\tau^* \geq n+1] V_\beta^*(X_n) \\ & \quad = \beta^n I[\tau^* \geq n+1]\left[R(X_n, f(X_n)) + \beta \sum_{y\in S} p_{x\,y}(f(X_n)V_\beta^*(y)\right] \\ & \quad = \beta^n R(X_n, f(X_n)) I[\tau^* > n] + I[\tau^* \geq n+1]\beta^{n+1} \sum_{y\in S} p_{x\,y}(f(X_n))V_\beta^*(y) \\ & \quad = E_x^{f^*}\left[\beta^n R(X_n, A_n) I[\tau^* > n] + \beta^{n+1} I[\tau^* \geq n+1] V_\beta^*(X_{n+1}) \middle| \mathcal{F}_n\right] \end{aligned} \tag{25}$$

where, using that $X_n \in S \setminus \mathcal{S}^*$ on the event $[\tau^* \geq n+1]$, the second equality is due to (21), and the Markov property was used in the last step. Therefore,

$$\begin{aligned} & E_x^{f^*}\left[\beta^n I[\tau^* \geq n+1] V_\beta^*(X_n)\right] \\ & \quad = E_x^{f^*}\left[\beta^n R(X_n, f(X_n)) I[\tau^* > n] + \beta^{n+1} I[\tau^* \geq n+1] V_\beta^*(X_{n+1})\right] \end{aligned}$$

a relation that together with (24) yields that

$$\begin{aligned}
&E_x^{f^*}\left[\beta^n I[\tau^* \geq n] V_\beta^*(X_n)\right] \\
&\quad = E_x^{f^*}\left[\beta^{\tau^*} I[\tau^* = n] G(X_{\tau^*})\right] + E_x^{f^*}\left[\beta^n R(X_n, f(X_n)) I[\tau^* > n]\right] \\
&\qquad\qquad + E_x^{f^*}\left[\beta^{n+1} I[\tau^* \geq n+1] V_\beta^*(X_{n+1})\right].
\end{aligned}$$

Combining this equality with the induction hypothesis, it follows that (23) holds with $n+1$ instead of n, completing the induction argument. To conclude, take the limit as $n \to \infty$ in (23) to obtain, *via* the bounded convergence theorem, that for each $x \in S \setminus \mathcal{S}^*$

$$\begin{aligned}
V_\beta^*(x) &= E_x^{f^*}\left[\sum_{t=0}^{\infty} \beta^t R(X_t, A_t) I[\tau^* > t] + \beta^{\tau^*} G(X_{\tau^*}) I[\tau^* < \infty]\right] \\
&= E_x^{f^*}\left[\sum_{t=0}^{\tau^*-1} \beta^t R(X_t, A_t) + \beta^{\tau^*} G(X_{\tau^*}) I[\tau^* < \infty]\right] \\
&= V_\beta(x, ; f^*, \tau^*);
\end{aligned}$$

see (1) and 2). This establishes that the equality (20) also holds when $x \in S \setminus \mathcal{S}^*$. □

Proof of Theorem 4.1. Let $(f^*, \tau^*) \in \mathbb{F} \times \mathcal{T}$ be as in (17) and (19).

It will be shown that for each strategy $\pi \in \mathcal{P}$ and $x \in S$,

$$V_\beta(x; \pi, \tau^*) \leq V_\beta(x; f^*, \tau^*). \tag{26}$$

To achieve this goal, first notice that if $x \in \mathcal{S}^*$ then the event $[\tau^* = 0]$ has probability 1 with respect to P_x^π and $P_x^{f^*}$, and in this case both sides of (26) are equal to $G(x)$, by (1). To verify the above inequality when x does not belong to $\mathcal{S}^*$, notice that the equilibrium equation (11) and (17) together imply that, for each $x \in S \setminus \mathcal{S}^*$ and $a \in A(x)$,

$$V_\beta^*(x) \geq R(x, a) + \beta \sum_{y \in S} p_{x\,y}(a) V_\beta^*(y), \tag{27}$$

and then, for every $\pi \in \mathcal{P}$ and $x \in S \setminus \mathcal{S}^*$,

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

where the second equality is due to the relation $P_x^\pi[\tau^* \geq 1] = 1$. From this point, combining the Markov property with (27), an induction argument along the lines used in the proof of Lemma 4.2 allows to establish that for each positive integer n and $\pi \in \mathcal{P}$,

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

taking the limit as n goes to ∞, *via* the bounded convergence theorem, it follows from (1) and (2) that, for each $x \in S \setminus \mathcal{S}^*$

$$\begin{aligned} V_\beta^*(x) &\geq E_x^\pi\left[\sum_{t=0}^\infty \beta^t R(X_t, A_t) I[\tau^* > t]\right] + E_x^\pi\left[\beta^{\tau^*} G(X_{\tau^*}) I[\tau^* < \infty]\right] \\ &= E_x^\pi\left[\sum_{t=0}^{\tau^*-1} \beta^t R(X_t, A_t)\right] + E_x^\pi\left[\beta^{\tau^*} G(X_{\tau^*}) I[\tau^* < \infty]\right] \\ &= V_\beta(x, \pi, \tau^*); \end{aligned}$$

see (1) and (2). This fact and Lemma 4.2 together yield that (26) also holds when $x \in S \setminus \mathcal{S}^*$.

To complete the proof of the theorem it will be shown that

$$V_\beta(x; f^*, \tau) \geq V_\beta(x; f^*, \tau^*), \quad \tau \in \mathcal{T}, \quad x \in S. \tag{28}$$

To achieve this goal, consider the reduced model

$$\hat{\mathcal{G}} = (S, A, \{\hat{A}(x)\}, R, G, P)$$

obtained from $\mathcal{G}$ by shrinking the action sets $A(x)$ to

$$\hat{A}(x) = \{f^*(x)\}, \quad x \in S,$$

and restricting the domain of $R(\cdot)$ and each $p_{x\,y}(\cdot)$ to $\hat{A}(x)$. For this new model, the corresponding class $\hat{\mathcal{P}}$ of strategies for player I is the singleton $\{f^*\}$, so that the (upper-) value function associated with $\hat{\mathcal{G}}$ is given by

$$\hat{V}_\beta^*(x) = \inf_{\tilde{\tau} \in \mathcal{T}} V_\beta(x; f^*, \tilde{\tau}), \quad x \in S, \tag{29}$$

an expression that is obtained from (4) by replacing $\mathcal{P}$ by $\hat{\mathcal{P}} = \{f^*\}$. By Theorem 3.1 applied to this reduced game $\hat{\mathcal{G}}$, the function $\hat{V}_\beta^*$ is characterized as the unique solution in $\mathcal{B}(S)$ of the equilibrium equation

$$\hat{V}_\beta^*(x) = \min\left\{G(x), \left[R(x, f^*(x)) + \beta \sum_{y \in S} p_{x\,y}(f^*(x)) \hat{V}_\beta^*(y)\right]\right\}, \quad x \in S.$$

Observe now that (11) and (19) together yield that the above equality is also valid if $\hat{V}_\beta^*$ is replaced by V_β^*, so that

$$\hat{V}_\beta^*(x) = V_\beta^*(x), \quad x \in S.$$

Combining this equality with (29), it follows that, for each $\tau \in \mathcal{T}$ and $x \in S$,

$$\begin{aligned} V_\beta(x; f^*, \tau) &\geq \inf_{\tilde{\tau} \in \mathcal{T}} V_\beta(x; f^*, \tilde{\tau}) \\ &= \hat{V}^*_\beta(x) \\ &= V^*_\beta(x), \end{aligned}$$

and then

$$V_\beta(x; f^*, \tau) \geq V_\beta(x; f^*, \tau^*),$$

by Lemma 4.2. This establishes (28), a relation that together with (26) yields that the pair (f^*, τ^*) is a Nash equilibrium; see Definition 2.1. □

5. APPROXIMATE EQUILIBRIA

The Nash equilibrium (f^*, τ^*) constructed above depends on the exact knowledge of the value function V^*_β. In this section it will be shown that, using sufficiently close approximations to V^*_β, it is possible to construct a pair of strategies which is 'nearly' a Nash equilibrium, an idea that is formally introduced below.

Definition 5.1. Let $\varepsilon > 0$ be arbitrary, and let $\mathcal{G}$ be the game described in Section 2. A pair $(\hat{f}, \hat{\tau}) \in \mathcal{P} \times \mathcal{T}$ is and ε-Nash equilibrium for $\mathcal{G}$ if

$$V_\beta(x; \pi, \hat{\tau}) < V_\beta(x; \hat{\pi}, \hat{\tau}) + \varepsilon, \quad x \in S, \quad \pi \in \mathcal{P},$$

and

$$V_\beta(x; \hat{\pi}, \tau) > V_\beta(x; \hat{\pi}, \hat{\tau}) - \varepsilon, \quad x \in S, \quad \tau \in \mathcal{T}.$$

Suppose that the players are willing to move from using their actual strategies only if the change represents an improvement of at least ε with respect to the current performance. In this case, when the strategies $\hat{\pi}$ and $\hat{\tau}$ actually used by the players form an ε-Nash equilibrium, if one player keeps on using his strategy, the other one will not have sufficiently strong incentives to change his behavior. The construction of ε-Nash equilibria is based on the following successive approximations scheme.

Definition 5.2. Let $W \in \mathcal{B}(S)$ be arbitrary but fixed.

(i) The sequence $\{W_n\} \subset \mathcal{B}(S)$ of successive approximations functions is defined as follows:

$$W_0 := W, \quad \text{and} \quad W_n = C[W_{n-1}], \quad n = 1, 2, 3, \ldots$$

(ii) For each positive integer n, set

$$\mathcal{S}^*_n := \{x \in S \mid W_{n+1} = G(x)\},$$

and define the pair $(f^*_n, \tau^*_n) \in \mathbb{F} \times \mathcal{T}$ by

$$\tau^*_n(\mathbf{h}) = \min\{t \geq 0 \mid x_t \in \mathcal{S}^*_n\}, \quad \mathbf{h} = (x_0, a_0, x_1, a_1, \ldots) \in \mathbb{H},$$

whereas $f_n^* \in \mathbb{F}$ is any policy satisfying

$$\begin{aligned} &R(x, f_n^*(x)) + \beta \sum_{y\in S} p_{x\,y}(f_n^*(x))W_n(y) \\ &\qquad = \sup_{a\in A(x)} \left[R(x,a) + \beta \sum_{y\in S} p_{x\,y}(a)W_n(y)\right], \quad x \in S; \end{aligned} \tag{30}$$

notice that the existence of such a policy f_n^* is ensured by Assumption 2.1.

The main contribution of this section is the following result.

Theorem 5.3. Let $\varepsilon > 0$ be fixed and, with the notation in Definition 5.2, let the positive integer n be such that

$$\varepsilon > 2\left[\frac{\beta^n \|C[W_0] - W_0\|}{1-\beta} + \beta^n \|C[W_0] - W_0\|\right]. \tag{31}$$

In this case, the pair (f_n^*, τ_n^*) is an ε-Nash equilibrium for the stopping game $\mathcal{G}$.

The proof of this theorem relies on the following auxiliary result.

Lemma 5.4. Let $\mathcal{G}$ be the stopping game described in Section 2, and consider the new game

$$\tilde{\mathcal{G}} = (S, A, \{A(x)\}_{x\in S}, \tilde{R}, \tilde{G}, P)$$

obtained from $\mathcal{G}$ by replacing the running and terminal rewards R and G by $\tilde{R} \in \mathcal{B}(\mathbb{K})$ and $\tilde{G} \in \mathcal{B}(S)$, respectively. For each $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$ let

$$\tilde{V}_\beta(x; \pi, \tau) := E_x^\pi \left[\sum_{t=0}^{\tau-1} \beta^t \tilde{R}(X_t, A_t) + \beta^\tau \tilde{G}(X_\tau)\right] \tag{32}$$

be the total expected reward of player I at state x under the pair (π, τ) in this new game. With this notation, the following assertions (i) and (ii) hold:

(i) For each pair $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$,

$$|\tilde{V}_\beta(x; \pi, \tau) - V_\beta(x; \pi, \tau)| \le \frac{\|\tilde{R} - R\|}{1-\beta} + \|\tilde{G} - G\|, \quad x \in S.$$

(ii) Let $(\tilde{f}, \tilde{\tau})$ be a Nash equilibrium for the game $\tilde{\mathcal{G}}$, and suppose that the real number ε satisfies that

$$\varepsilon > 2\left[\frac{\|\tilde{R} - R\|}{1-\beta} + \|\tilde{G} - G\|\right]. \tag{33}$$

In this case, the pair $(\tilde{f}, \tilde{\tau})$ is an ε-Nash equilibrium for the original game $\mathcal{G}$.

Proof. The first assertion follows combining (1) and (32). As for the second claim, first notice that when ε satisfies (33), part (i) yields that the inequality

$$|\tilde{V}_\beta(x;\pi,\tau) - V_\beta(x;\pi,\tau)| < \frac{\varepsilon}{2} \tag{34}$$

is always valid, and let $(\tilde{f},\tilde{\tau})$ be a Nash equilibrium for the game $\tilde{\mathcal{G}}$, that is,

$$\tilde{V}_\beta(x;\pi,\tilde{\tau}) \le \tilde{V}_\beta(x;\tilde{\pi},\tilde{\tau}), \quad x \in S, \quad \pi \in \mathcal{P},$$

and

$$\tilde{V}_\beta(x;\tilde{\pi},\tau) \ge \tilde{V}_\beta(x;\tilde{\pi},\tilde{\tau}), \quad x \in S, \quad \tau \in \mathcal{T}.$$

Combining these two last relations with (34), it follows that

$$V_\beta(x;\pi,\tilde{\tau}) < V_\beta(x;\tilde{\pi},\tilde{\tau}) + \varepsilon, \quad x \in S, \quad \pi \in \mathcal{P},$$

and

$$V_\beta(x;\tilde{\pi},\tau) > V_\beta(x;\tilde{\pi},\tilde{\tau}) - \varepsilon, \quad x \in S, \quad \tau \in \mathcal{T},$$

and then, by Definition 5.1, the pair $(\tilde{\pi},\tilde{\tau})$ is an ε-Nash equilibrium for the original game $\mathcal{G}$. □

Proof of Theorem 5.3. Keeping in mind the specifications of the operator C and the sequence $\{W_n\}$ as in (6) and Definition 5.2, respectively, notice that (8) yields that

$$\|W_{n+1} - W_n\| \le \beta^n \|W_1 - W_0\| = \beta^n \|C[W_0] - W_0\|, \tag{35}$$

whereas the equality $W_{n+1} = C[W_n]$ can be explicitly written as

$$W_{n+1}(x) = \min\left\{ G(x), \sup_{a\in A(x)} \left[R(x,a) + \beta \sum_{y\in S} p_{x\,y}(a) W_n(y) \right] \right\}, \quad x \in S. \tag{36}$$

Now, define the new reward functions $R_n \in \mathcal{B}(\mathbb{K})$ and $G_n \in \mathcal{B}(S)$ by

$$\begin{aligned} R_n(x,a) &:= R(x,a) - [W_{n+1}(x) - W_n(x)], \quad (x,a) \in \mathbb{K}, \\ G_n(x) &:= G(x) - [W_{n+1}(x) - W_n(x)], \quad x \in S, \end{aligned} \tag{37}$$

and notice that (35) yields that

$$\|R - R_n\| \le \beta^n \|C[W_0] - W_0\| \quad \text{and} \quad \|G - G_n\| \le \beta^n \|C[W_0] - W_0\|,$$

so that the condition (31) yields that

$$\varepsilon > 2\left[\frac{\|R - R_n\|}{1-\beta} + \|G - G_n\| \right]. \tag{38}$$

Consider now the new stopping game $\mathcal{G}_n = (S, A, \{A(x)\}_{x\in S}, R_n, G_n, P)$, and let $V^*_{n\,\beta}$ be the value function corresponding to $\mathcal{G}_n$. Observing that (39) and (37) together lead to

$$W_n(x) = \min\left\{ G_n(x), \sup_{a\in A(x)} \left[R_n(x,a) + \beta \sum_{y\in S} p_{x\,y}(a) W_n(y) \right] \right\}, \quad x \in S, \tag{39}$$

an application of Theorem 3.1(iii) to the game $\mathcal{G}_n$ yields that $W_n = V^*_{n\,\beta}$. Next, observe that, (37) shows that the specification of the set $\mathcal{S}^*_n$ in Definition 5.2 is equivalent to

$$\mathcal{S}^*_n = \{x \in S \,|\, G_n(x) = V^*_{n\,\beta}(x)\},$$

so that τ^*_n in (5.2) is the first arrival time to the set where the value function and the terminal reward of the game $\mathcal{G}_n$ coincide. On the other hand, (30) and (37) together with the equality $W_n = V^*_{n\,\beta}$ yield that, for each $x \in S$,

$$R_n(x, f^*_n(x)) + \beta \sum_{y\in S} p_{x\,y}(f^*_n(x)) V^*_{n\,\beta}(y) = \sup_{a\in A(x)} \left[R_n(x,a) + \beta \sum_{y\in S} p_{x\,y}(a) V^*_{n\,\beta}(y) \right],$$

and then an application of Theorem 4.1 to the game $\mathcal{G}_n$ yields that the pair $(\tau^*_n, f^*_n) \in \mathcal{P} \times \mathcal{T}$ is a Nash equilibrium for $\mathcal{G}_n$. Combining this fact with (38), an application of Lemma 5.4 with $\mathcal{G}_n$ instead of $\tilde{\mathcal{G}}$ yields that the pair (τ^*_n, f^*_n) is an ε-Nash equilibrium for the original stopping game $\mathcal{G}$. □

6. AN EXAMPLE: APPROXIMATING A HEDGING PROBLEM

This section presents an example motivated by the optimal hedging problem in mathematical finance, which can be formulated as a Markov stopping game of the form described in the previous sections. The analysis also points out some positive aspects of this approach to approximate other kind of problems. The financial market is the same as in Bielecki et al. [3], and is formulated as follows: Let X_t be a finite state Markov chain with transition matrix $Q = [Q_{x,y}]$ representing external economic factors taking values in S that are involved in the evolution of the relative prices Z of the risky asset via the conditional probability distribution $\nu(x, y, dz)$, with compact support. Besides the risky asset, there is a bank account paying a constant interest rate r. The initial capital of the investor is 1, and his admissible actions belong to some compact set of $\mathbb{R}$, whose elements represent the proportion of wealth invested in the risky asset. The rest of the capital is invested in the bank account. An admissible strategy of the investor is a sequence of stochastic kernels $\pi = \{\pi_t\}$ as described in Section 2, and the value of the portfolio evolves according to

$$V_{t+1} = V_t[e^r + \pi_t(Z_{t+1} - e^r)].$$

Given a function $h : S \to \mathbb{R}$, consider the following reward

$$\log V_\tau - h(X_\tau),$$

where τ is a stopping time adapted to the information received from the evolution of the economic factors and the decisions of the investor. This kind of rewards are motivated by the optimal hedging problem of American options in finance. Here player I is the investor, with action space given by A, and player II is the owner of the American option, with the right to exercise the option at any stopping time τ. In order to work in a simple setting, the option was written in terms of X_t, but it can be expressed in terms of the price of the risky asset. Using the notation in the previous section, define the upper value function of the game as

$$V^*(x) = \inf_{\tau \in \mathcal{T}} \left[\sup_{\pi \in \mathcal{P}} E_x^\pi \left[\log V_\tau - h(X_\tau) \right] \right], \quad x \in S, \tag{40}$$

where the following are enforced:

(a) For each $a \in A$, $e^r + a(Z - e^r) > 0$, and

(b) For each $x, y \in S$ and $a \in A$, the integral $\int \log \left[e^r + a(z - e^r) \right] \nu(x, y, \mathrm{d}z)$ is finite.

Defining the reward function

$$R(x, a) = \sum_{y \in S} Q_{x,y} \int \log \left[e^r + a(z - e^r) \right] \nu(x, y, \mathrm{d}z),$$

after some calculations involving conditional expectations along the lines in Bielecki et al. [3], it is possible to write the functional in (40) as

$$E_x^\pi \left[\log V_\tau - h(X_\tau) \right] = E_x^\pi \left[\sum_{t=0}^{\tau - 1} R(X_t, A_t) - h(X_\tau) \right], \quad x \in S,$$

an expression that, except for the absence of the discount factor, casts with (1). In fact, for β near to 1, the results on the discounted criterion obtained in this note, can be seen as an approximation to the above hedging problem.

ACKNOWLEDGEMENT

The authors are grateful to the reviewers for their careful reading of the original manuscript and their helpful suggestions to improve the paper.
This work was supported in part by the PSF Organization under Grant No. 030711, and in part by CONACYT under Grant No. 61423.

(Received December 12, 2011)

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* (2000), pp. 213–221, Birkhauser, Boston.

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

[3] 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. Mathe. Methods Oper. Res. *50* (1999), 167–188.

[4] E. B. Dynkin: The optimum choice for the instance for stopping Markov process. Soviet. Math. Dokl. *4* (1963), 627–629.

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

[6] G. Peskir: On the American option problem. Math. Finance *15* (2010), 169–181.

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

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

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

[10] 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.

[11] K. Sladký: Risk-sensitive Ramsey Growth model. In: Proc. of 28th International Conference on Mathematical Methods in Economics (M. Houda and J. Friebelová, eds.) České Budějovice 2010.

[12] L. S. Shapley: Stochastic games. Proc. Nat. Acad. Sci. U.S.A. *39* (1953), 1095–1100.

[13] J. van der Wal: Discounted Markov games: Successive approximation and stopping times. Internat. J. Game Theory *6* (1977), 11–22.

[14] J. van der Wal: Discounted Markov games: Generalized policy iteration method. J. Optim. Theory Appl. *25* (1978), 125–138.

[15] D. J. White: Real applications of Markov decision processes. Interfaces *15* (1985), 73–83.

[16] D. J. White: Further real applications of Markov decision processes. Interfaces *18* (1988), 55–61.

[17] L. E. Zachrisson: Markov games. In: Advances in Game Theory (M. Dresher, L. S.Shapley and A. W. Tucker, eds.), Princeton Univ. Press, Princeton 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: rcavazos@uaaan.mx

Daniel Hernández-Hernández, Centro de Investigación en Matemáticas, Apartado Postal 402, Guanajuato, GTO 36000. México.
e-mail: dher@cimat.mx