KYBERNETIKA — VOLUME *60* (2024), NUMBER 1, PAGES 1–18

DENUMERABLE MARKOV STOPPING GAMES WITH RISK-SENSITIVE TOTAL REWARD CRITERION

Manuel A. Torres-Gomar, Rolando Cavazos-Cadena and
Hugo Cruz-Suárez

This paper studies Markov stopping games with two players on a denumerable state space. At each decision time player II has two actions: to stop the game paying a terminal reward to player I, or to let the system to continue it evolution. In this latter case, player I selects an action affecting the transitions and charges a running reward to player II. The performance of each pair of strategies is measured by the risk-sensitive total expected reward of player I. Under mild continuity and compactness conditions on the components of the model, it is proved that the value of the game satisfies an equilibrium equation, and the existence of a Nash equilibrium is established.

Keywords: monotone operator, fixed point, equilibrium equation, Nash equilibrium, hitting time, bounded rewards

Classification: 91A10, 91A15

1. INTRODUCTION

This note concerns with discrete-time Markov stopping games evolving on a denumerable state space. The system is directed by two players, and at each decision time player II always can choose between two possible actions, namely, to stop the game, or to let the system to continue its evolution. In this latter case, player I selects an action affecting the transitions and charges a running reward to player II, whereas if the game is stopped, player II pays a terminal reward to player I. It is supposed that player I has a nonull and constant risk sensitivity coefficient, and tries to maximize his total (risk-sensitive) expected reward, whereas player II tries to minimize it. Besides mild continuity conditions, the main structural conditions on the game concern with the reward structure, namely, it is supposed that the terminal reward function is nonnegative, and that the running reward is always larger than a positive constant. In this framework, the main problems analyzed in this note are as follows:

- to establish an equilibrium equation characterizing the value function of the game, and
- to establish the existence of a Nash equilibrium.

DOI: 10.14736/kyb-2024-1-0001

In a risk-neutral context, these problems have been recently studied in Martínez-Cortés (2021) [20], where Markov stopping games with total reward criterion on a finite state space were analyzed under the assumption that the system has an absorbing state, a context that generalizes the discounted framework in Cavazos-Cadena and Hernández–Hernández (2012) [8]. On the other hand, in Cavazos–Cadena et al. (2021) [9] the risk-neutral total reward criterion was studied under the assumption that the state space is denumerable and the system has an absorbing state, whereas in Cavazos–Cadena et al. (2021) [10] this last condition was replaced by the assumption that the Markov chain induced by any stationary policy of player I is communicating and has an invariant distribution. More recently, in a risk-sensitive framework the problems posed above were studied in López-Rivero et al. (2022) [19] under the condition that the system has an absorbing state which can be reached regardless of the initial point. The results presented in this paper extend those reported in Cavazos–Cadena and Hernández-Hernández (2012) [8], where the total expected discounted reward is considered as a performance criterion. The main difference between the results presented in this paper and those already available in the literature can be summarized as follows: the conclusions obtained in this note involve only mild restrictions on the reward structure but, in contrast with other related works, no restriction is imposed on the transition law.

A starting benchmark in the study of discrete-time Markov models with risk-sensitive criteria was settled in Howard and Matheson (1972) [16], where controlled Markov chains on a finite state space were analyzed and, under communication conditions, the optimal risk-sensitive average cost was characterized in terms of an optimality equation. Those basic results have been extended to the case of a general transition structure in Cavazos-Cadena and Hernández-Hernández (2006) [7] for uncontrolled models, and for controlled models in Alanís Durán and Cavazos-Cadena (2012) [1]. Controlled Markov decision processes with finite or denumerable state-space endowed with risk-sensitive criteria have been studied, for instance, in Bäuerle and Rieder (2014) [3], Borkar and Meyn (2002) [6], Denardo and Rothblum (2006) [11], Sladký (2009, 2010, 2018) ([23, 24, 25]) whereas Markov decision processes on a general state space are analyzed, for instance, in Di Masi and Stettner (1999, 2000, 2007) ([12, 13, 14]) or Jaśkiewicz (2007) [17]. Applications of risk-sensitive criteria are presented, for example, in the area of mathematical finance (Bäuerle and Rieder 2011 [2], Bielecki et al. 1999 [5], Pitera and Stettner 2016 [21], Stettner 1999 [26]), and in large deviations (Balaji and Meyn 2000 [4], Kontoyiannis and Meyn, 2003 [18]).

The remainder of the paper is organized as follows: In Section 2 the Markov stopping game is formally defined, the performance criterion is formulated, the concept of Nash equilibrium is introduced and the main structural conditions are stated as Assumption 2.1. Next, the main conclusions of the paper are presented as Theorem 3.3, a result that is proved in Section 5 after the necessary preliminaries given in Section 4. The approach to achieve this goal is to apply results referent to stopping times as well as dynamic programming techniques. Finally, the paper concludes with some brief comments in Section 6.

Notation: Given a nonempty set $\mathbb{M}$, the Banach space $\mathcal{B}(\mathbb{M})$ consist of all continuous functions $R\colon \mathbb{M} \to \mathbb{R}$ whose supremum norm $\|R\|$ is finite, where $\|R\| := \sup_{k\in\mathbb{M}} |R(k)|$, whereas $\mathbb{N}$ stands for the set of nonnegative integers. The indicator function of an

event A is denoted by $I[A]$ and, even without explicit mention, all relations involving conditional expectations are valid with probability 1 with respect to the underlying probability measure. The minimum of the empty set is ∞ and, finally, the following convention concerning summations will be used:

$$\sum_{t=n}^{m} a_t := 0, \quad m < n. \tag{1}$$

2. THE MODEL

In this section the dynamic model studied in the paper is formally introduced. A Markov stopping game with two players labeled I and II is given by $\mathcal{G} = (S, A, \{A(x)\}_{x\in S}, R, G, P)$, a mathematical structure whose components have the following meaning: The (nonempty) 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. The next components, $R \in \mathcal{B}(\mathbb{K})$ and $G \in \mathcal{B}(S)$ are the running and terminal reward functions, respectively, where $\mathbb{K} := \{(x, a) \,|\, a \in A(x), x \in S\}$ is the class $\mathbb{K}$ of admissible pairs. 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 $(x, a) \in \mathbb{K}$. The evolution of the dynamic system is as follows: Each player observes the state of the system at each time $t \in \mathbb{N}$, 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 to let the system *to continue* its evolution. In this latter case, using the observed state $X_t = x$ as well the record of previous states and actions, player I selects and applies an action (control) $A_t = a \in A(x)$, an intervention which has two consequences: (a) player I gets a reward $R(x, a)$ from player II, and (b) the system jumps to $X_{t+1} = y \in S$ with probability $p_{x\,y}(a)$; this is the Markov property of the decision process. The following structural assumption will be enforced.

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) $G(x) \geq 0$ for ech $x \in S$.

(iv) There exists $\delta > 0$ such that $R(x, a) \geq \delta$ for every $(x, a) \in \mathbb{K}$.

Decision Strategies. Consider the model $\mathcal{G}$ defined above and, for each $t = 0, 1, ...$ define the space $\mathbb{H}_t$ of possible histories up to time t as $\mathbb{H}_0 := S$ and $\mathbb{H}_t := \mathbb{K}^t \times S$ when $t > 0$. A generic element of $\mathbb{H}_t$ is a vector of the form $\mathbf{h}_t = (x_0, a_0, \ldots, x_i, a_i, \ldots, x_t)$ where $a_i \in A(x_i)$ and $x_i \in S$. Then a control policy is a sequence $\pi = \{\pi_t\}$ of stochastic kernels, that is, 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 the mapping $\mathbf{h}_t \mapsto \pi_t(B|\mathbf{h}_t)$, $\mathbf{h}_t \in \mathbb{H}_t$, is Borel measurable for each Borel subset $B \subset A$. The class of all policies constitutes the family of *admissible strategies for player I* and is denoted by $\mathcal{P}$. Thus, 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 the policy $\pi \in \mathcal{P}$ and the initial state $X_0 = x$, the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined (Hernández-Lerma 1988 [15], Puterman 1994 [22]); such distribution is denoted by P_x^π which is defined on the Borel σ-field of the space $\mathbb{H} := \prod_{t=0}^{\infty} \mathbb{K}$, and 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 $f\colon S \to A$ 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 $\pi_t(\{f(x_t)\}|\mathbf{h}_t) = 1$, and in this case π and f are naturally identified, a convention allowing to write $\mathbb{F} \subset \mathcal{P}$. Next, let the σ-algebra $\mathcal{F}_t$ be defined by

$$\mathcal{F}_t\colon = \sigma(X_0, A_0, \ldots, X_{t-1}, A_{t-1}, X_t), \quad t \in \mathbb{N}, \tag{2}$$

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

Exponential Utility. Throughout the remainder it is supposed that player I has a constant risk-sensitivity coefficient $\lambda \neq 0$, which means that a random reward Y is assessed via the expectation of $U_\lambda(Y)$, where the utility function $U_\lambda\colon \mathbb{R} \to \mathbb{R}$ is defined by

$$U_\lambda(x) := \operatorname{sign}(\lambda) e^{\lambda x}, \quad x \in S; \tag{3}$$

observe that $U_\lambda(\cdot)$ is a strictly increasing function and that the following relation holds

$$U_\lambda(x + y) = e^{\lambda x} U_\lambda(y), \quad x, y \in \mathbb{R}; \tag{4}$$

as usual, set $U_\lambda(\infty) := \infty$ if $\lambda > 0$ and $U_\lambda(\infty)\colon = 0$ if $\lambda < 0$. When choosing between two random rewards W and Y, player I prefers Y if $E[U_\lambda(W)] < E[U_\lambda(Y)]$, and is indifferent between both rewards when $E[U_\lambda(W)] = E[U_\lambda(Y)]$. The certainty equivalent of Y (with respect to U_λ) is the constant $\mathcal{E}_\lambda(Y) \in \mathbb{R} \cup \{-\infty, \infty\}$ satisfying $U_\lambda(\mathcal{E}_\lambda(Y)) = E[U_\lambda(Y)]$, so that player I is indifferent between receiving a random reward Y or the corresponding certainty equivalent $\mathcal{E}_\lambda(Y)$; observe that

$$\mathcal{E}_\lambda(Y) := \log(E[e^{\lambda Y}])/\lambda. \tag{5}$$

Performance Criterion. Given the initial state $X_0 = x \in S$, suppose that players I and II drive the system using strategies $\pi \in \mathcal{P}$ and $\tau \in \mathcal{T}$, respectively. The total (random) reward obtained by player I until the system is halted at time τ by player II is given by

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

and the corresponding certainty equivalent is the *performance index* $V_\lambda(x; \pi, \tau)$ associated with the pair $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$ at state $x \in S$, that is,

$$V_\lambda(x; \pi, \tau)\colon = \frac{1}{\lambda} \log \left(E_x^\pi \left[I[\tau < \infty] e^{\lambda(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau))} + I[\tau = \infty] e^{\lambda \sum_{t=0}^{\infty} R(X_t, A_t)} \right] \right); \tag{6}$$

see (5). Observe that the nonnegativity of R and G yield that

$$V_\lambda(x;\pi,\tau) \geq 0, \tag{7}$$

and that (6) is equivalent to

$$U_\lambda(V_\lambda(x;\pi,\tau)) = E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) \right) I[\tau < \infty] \right] + E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{\infty} R(X_t, A_t) \right) I[\tau = \infty] \right]. \tag{8}$$

Given that player II employs the strategy τ and the initial state is x, player I tries to use a policy $\tilde{\pi}$ such that $V_\lambda(x;\tilde{\pi},\tau) = \sup_{\pi\in\mathcal{P}} V_\lambda(x;\pi,\tau)$, which is a function of x and τ, say $\varphi(x;\tau)$. In this work it is supposed that the *main objective of player* II *is to minimize the utility gained by player* I, and then player II will try to use a stopping time $\tilde{\tau}$ such that $\varphi(x;\tilde{\tau})$ is as close as possible to $\inf_{\tau\in\mathcal{T}} \varphi(x;\tau)$, which is the (*upper-*)*value* function of the game and is explicitly determined by

$$V_\lambda^*(x) \colon = \inf_{\tau\in\mathcal{T}} \left[\sup_{\pi\in\mathcal{P}} V_\lambda(x;\pi,\tau) \right], \quad x \in S. \tag{9}$$

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

$$V_{\lambda,*}(x) \colon = \sup_{\pi\in\mathcal{P}} \left[\inf_{\tau\in\mathcal{T}} V_\lambda(x;\pi,\tau) \right], \quad x \in S. \tag{10}$$

Since $\sup_{\pi\in\mathcal{P}} V_\lambda(x;\pi,\tau) \geq V_\lambda(x;\pi,\tau) \geq \inf_{\tau\in\mathcal{T}} V_\lambda(x;\pi,\tau)$, these definitions immediately lead to

$$V_\lambda^*(\cdot) \geq V_{\lambda,*}(\cdot). \tag{11}$$

The remainder of the paper is dedicated to establish 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 $x \in S$ the following relation holds:

$$V_\lambda(x;\pi,\tau^*) \leq V_\lambda(x;\pi^*,\tau^*) \leq V_\lambda(x;\pi^*,\tau), \quad \pi\in\mathcal{P}, \quad \tau\in\mathcal{T}. \tag{12}$$

It follows that if the strategies π^* and τ^* actually used by players I and II form a Nash equilibrium, then if player II keeps on using strategy τ^*, then the first inequality in (12) yields that player I does not have any incentive to switch to other policy. Similarly, the second inequality in (12) implies that if player I keeps on using π^*, then player II does not have any motivation to change the strategy τ^* in use. Also, note that if (π^*,τ^*) is a Nash equilibrium, then (12) implies that

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

where the left- and right-most inequalities are due to (9) and (10), respectively, so that via (11), it follows that the upper and lower value functions are equal and coincide with $V_\lambda(\cdot;\pi^*,\tau^*)$, and in this case this function is the value of the game. As already mentioned, the main objective of the paper is to establish the existence of a Nash equilibrium, and the result in this direction will stated in the following section.

3. MAIN THEOREM

In this section the main result of this note is stated. To this end the notation in [19] will be followed closely.

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

$$[\![0,G]\!] := \{h \in \mathcal{C}(S) \,|\, 0 \leq h(x) \leq G(x)\}, \tag{13}$$

whereas the operator $T_\lambda : [\![0,G]\!] \to [\![0,G]\!]$ is defined as follows: For each $W \in [\![0,G]\!]$ and $x \in S$,

$$T_\lambda[W](x) := U_\lambda^{-1}\left(\min\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)}\sum_{y\in S} p_{x,y}(a)U_\lambda(R(x,a)+W(y))\right\}\right). \tag{14}$$

Using that R and G are nonnegative and that $U_\lambda(\cdot)$ is increasing, it is not difficult to prove that T_λ transforms $[\![0,G]\!]$ into itself, as well as to show that the following monotonicity property holds:

$$W, W_1 \in [\![0,G]\!] \text{ and } W \leq W_1 \implies T_\lambda[W] \leq T_\lambda[W_1]. \tag{15}$$

A most important property of T_λ, which was established in López–Rivero et al. (2022) [19], states that T_λ is continuous when $[\![0,G]\!]$ is endowed with the topology of pointwise-convergence.

Lemma 3.2. Under Assumption 2.1, suppose that the sequence $\{V_n\} \subset [\![0,G]\!]$ converges pointwise to V, that is,

$$\lim_{n\to\infty} V_n(y) = V(y), \quad y \in S.$$

In this case, $V \in [\![0,G]\!]$ and

$$\lim_{n\to\infty} T_\lambda[V_n](x) = T_\lambda[V](x), \quad x \in S.$$

This result was established as Theorem 4.1 in [19], where the argument relies only on the properties stated in Assumption 2.1. Next, given $\lambda \neq 0$, define the sequence $\{W_{\lambda,n} : S \to \mathbb{R}\}_{n\in\mathbb{N}}$ as follows:

$$W_{\lambda,0} = 0, \qquad W_{\lambda,n+1} = T_\lambda[W_{\lambda,n}], \quad n \in \mathbb{N}. \tag{16}$$

Since R and G are nonnegative, from (13) – (15) it is not difficult to see that

$$0 \leq W_{\lambda,n}(x) \leq W_{\lambda,n+1}(x) \leq G(x), \quad x \in S, \quad n \in \mathbb{N}, \tag{17}$$

so that, for each $x \in S$, $\{W_{\lambda,n}(x)\}$ converges to a point in $[0, G(x)]$. Set

$$W^*_\lambda(x) = \lim_{n\to\infty} W_{\lambda,n}(x) \in [0, G(x)], \quad x \in S. \tag{18}$$

Taking the limit as n goes to ∞ in both sides of the second equality in (16), via Lemma 3.2 it follows that

$$W^*_\lambda = T_\lambda[W^*_\lambda], \tag{19}$$

that is W^*_λ is a fixed point of T_λ. Notice that via Definition 3.1, the above display is equivalent to

$$U_\lambda(W^*_\lambda(x)) = \min\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a) + W^*_\lambda(y))\right\}, \quad x \in S. \tag{20}$$

Moreover, using Assumption 2.1, the inclusion $W^*_\lambda \in [\![0, G]\!]$ and the boundedness of G together imply that there exists a policy $f^* \in \mathbb{F}$ such that,

$$\begin{aligned}&\sum_{y\in S} p_{x,y}(f^*(x)) U_\lambda(R(x, f^*(x)) + W^*_\lambda(y)) \\ &\qquad = \sup_{a\in A(x)} \left[\sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a) + W^*_\lambda(y))\right], \quad x \in S. \end{aligned} \tag{21}$$

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

$$S^* :\ = \{x \in S \,|\, W^*_\lambda(x) = G(x)\}, \tag{22}$$

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

$$\tau^* :\ = \min\{n \in \mathbb{N} \,|\, X_n \in S^*\}, \tag{23}$$

so that τ^* is a stopping time with respect to the filtration $\{\mathcal{F}_t\}$ in (2), and then τ^* belongs to the space $\mathcal{T}$ of admissible strategies for player II. With this notation, the main conclusion of this note can be stated as follows.

Theorem 3.3. Under Assumption 2.1 the following assertions (i) – (iii) hold.
(i) For every $x \in S$, and $\pi \in \mathcal{P}$

$$V_\lambda(x; \pi, \tau^*) \le W^*_\lambda(x).$$

(ii) For every $x \in S$, and $\tau \in \mathcal{T}$,

$$V_\lambda(x; f^*, \tau) \ge W^*_\lambda(x).$$

(iii) $W^*_\lambda(\cdot) = V_\lambda(\cdot; f^*, \tau^*)$, and the pair $(f^*, \tau^*) \in \mathbb{F} \times \mathcal{T}$ is a Nash equilibrium .

This theorem will be established in the following section. Throughout the remainder Assumption 2.1 is enforced.

4. PROOF OF THE MAIN RESULT

This section contains the main tools that will be used to establish Theorem 3.3. The two main results are stated separately in two lemmas below using the notation in (16) – (23).

Lemma 4.1. For every $x \in S$

$$V_\lambda(x;\pi,\tau^*) \leq W_\lambda^*(x), \quad \pi \in \mathcal{P}. \tag{24}$$

Proof. For each $x \in S \setminus S^*$, note that $W_\lambda^*(x) < G(x)$, by (22). Thus, using (20) it follows that

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) &= \sup_{\tilde{a}\in A(x)} \sum_{y\in S} p_{x,y}(\tilde{a}) U_\lambda(R(x,\tilde{a}) + W_\lambda^*(y)) \\ &\geq \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a) + W_\lambda^*(y)), \quad x \in S\setminus S^*, \quad a \in A(x). \end{aligned}$$

Now let $\pi = \{\pi_k\} \in \mathcal{P}$ be arbitrary. Integrating with respect to $\pi_0(\cdot|x)$ it follows that

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) &\geq \sum_{y\in S} \int_{A(x)} p_{x,y}(a) U_\lambda(R(x,a) + W_\lambda^*(y))\, \pi_0(\mathrm{d}a|x) \\ &= E_x^\pi \left[U_\lambda(R(X_0,A_0) + W_\lambda^*(X_1)) \right], \quad \pi \in \mathcal{P}, \quad x \in S\setminus S^*, \end{aligned}$$

and via the Markov property it follows that

$$\begin{aligned} &U_\lambda(W_\lambda^*(X_n)) \\ &\quad \geq E_{X_n}^\pi \left[U_\lambda(R(X_n,A_n) + W_\lambda^*(X_{n+1})) \right] \\ &\quad = E_x^\pi \left[U_\lambda(R(X_n,A_n) + W_\lambda^*(X_{n+1})) | \, \mathcal{F}_n \right] \text{ on the event } [X_n \in S\setminus S^*], \quad n \in \mathbb{N}. \end{aligned} \tag{25}$$

Next, it will be shown that for every $x \in S$ and $\pi \in \mathcal{P}$,

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) \geq \sum_{k=0}^{n} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t,A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \\ + E_x^\pi \left[U_\lambda \left(\sum_{k=0}^{n-1} R(X_k,A_k) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right], \quad n \in \mathbb{N}. \end{aligned} \tag{26}$$

To establish this assertion consider the following two cases:

Case 1: $x \in S^*$.

In this case, using that $P_x^\pi[X_0 = x] = 1$ it follows from (23) that $\tau_* = 0$ P_x^π-a. s., and then the right-hand-side of (26) simplifies to

$$E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{0-1} R(X_t,A_t) + W_\lambda^*(X_0) \right) I[\tau^* = 0] \right] = E_x^\pi \left[U_\lambda \left(W_\lambda^*(X_0) \right) \right] = U_\lambda(W_\lambda^*(x))$$

so that (26) is equivalent to the true statement $U_\lambda(W_\lambda^*(x)) \geq U_\lambda(W_\lambda^*(x))$.

Case 2: $x \in S \setminus S^*$.

In this context, it will be shown, by induction, that (26) is valid for every $\pi \in \mathcal{P}$. Observing that $P_x^\pi[\tau^* = 0] = 0 = 1 - P_x^\pi[\tau^* > 0]$, by (23), for $n = 0$ (26) simplifies to

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) &\geq E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{0-1} R(X_k, A_k) + W_\lambda^*(X_0)\right) I[\tau^* > 0]\right] \\ &= E_x^\pi\left[U_\lambda\left(W_\lambda^*(X_0)\right) I[\tau^* > 0]\right] = U_\lambda(W_\lambda^*(x)) \end{aligned}$$

showing that the claim is valid for $n = 0$. Suppose now that (26) holds for some $n \in \mathbb{N}$, and observe that

(a) $\sum_{k=0}^{n-1} R(X_k, A_k) + W_\lambda^*(X_n)$ and $I[\tau^* > n]$ are $\mathcal{F}_n$-measurable, by (2), as well as

(b) $I[\tau^* > n] = [X_k \notin S^*, 0 \leq k \leq n] \subset [X_n \notin S^*]$. Thus

$$\begin{aligned} &E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{n-1} R(X_k, A_k) + W_\lambda^*(X_n)\right) I[\tau^* > n]\middle| \mathcal{F}_n\right] \\ &\quad = U_\lambda\left(\sum_{k=0}^{n-1} R(X_k, A_k) + W_\lambda^*(X_n)\right) I[\tau^* > n] \quad \text{(by (a))} \\ &\quad = e^{\lambda\sum_{k=0}^{n-1} R(X_k,A_k)} I[\tau^* > n] U_\lambda\left(W_\lambda^*(X_n)\right) \quad \text{(by (4))} \\ &\quad \geq e^{\lambda\sum_{k=0}^{n-1} R(X_k,A_k)} I[\tau^* > n] E_x^\pi\left[U_\lambda(R(X_n, A_n) + W_\lambda^*(X_{n+1}))\middle| \mathcal{F}_n\right] \\ &\quad = E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{n} R(X_k, A_k) + W_\lambda^*(X_{n+1})\right) I[\tau^* > n]\middle| \mathcal{F}_n\right] \end{aligned}$$

where the inequality was obtained combining (25) with property (b), and (4) together with property (a) were used in the last step. Thus,

$$\begin{aligned} &E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{n-1} R(X_k, A_k) + W_\lambda^*(X_n)\right) I[\tau^* > n]\right] \\ &\quad \geq E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{n} R(X_k, A_k) + W_\lambda^*(X_{n+1})\right) I[\tau^* > n]\right] \\ &\quad = E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{n} R(X_k, A_k) + W_\lambda^*(X_{n+1})\right) I[\tau^* = n+1]\right] \\ &\qquad + E_x^\pi\left[U_\lambda\left(\sum_{k=0}^{n} R(X_k, A_k) + W_\lambda^*(X_{n+1})\right) I[\tau^* > n+1]\right] \end{aligned}$$

and combining this relation with the induction hypothesis it follows that (26) holds with $n + 1$ instead of n, completing the induction argument, so that (26) is also valid when

$x \in S \setminus S^*$. To conclude, using that W_λ^* is nonnegative, notice that (26) implies that

$$U_\lambda(W_\lambda^*(x)) \geq \sum_{k=0}^{n} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \tag{27}$$
$$+ E_x^\pi \left[U_\lambda \left(\sum_{k=0}^{n-1} R(X_k, A_k)) \right) I[\tau^* > n] \right], \quad x \in S, \quad \pi \in \mathcal{P}.$$

Observe now that, by monotone convergence, as $n \to \infty$

$$\sum_{k=0}^{n} E_x^\pi \left[e^{\lambda\left(\sum_{t=0}^{k-1} R(X_t,A_t)+W_\lambda^*(X_k)\right)} I[\tau^* = k] \right] \to \sum_{k=0}^{\infty} E_x^\pi \left[e^{\lambda\left(\sum_{t=0}^{k-1} R(X_t,A_t)+W_\lambda^*(X_k)\right)} I[\tau^* = k] \right]$$
$$= \sum_{k=0}^{\infty} E_x^\pi \left[e^{\lambda\left(\sum_{t=0}^{k-1} R(X_t,A_t)+W_\lambda^*(X_{\tau^*})\right)} I[\tau^* = k] \right]$$
$$= \sum_{k=0}^{\infty} E_x^\pi \left[e^{\lambda\left(\sum_{t=0}^{k-1} R(X_t,A_t)+G_\lambda^*(X_{\tau^*})\right)} I[\tau^* = k] \right]$$
$$= E_x^\pi \left[e^{\lambda\left(\sum_{t=0}^{\tau^*-1} R(X_t,A_t)+G_\lambda^*(X_{\tau^*})\right)} I[\tau^* < \infty] \right],$$

where the the second equality is due to the fact that $W_\lambda^*(X_{\tau^*}) = G(X_{\tau^*})$ on the event $[\tau^* < \infty]$. Via (3) the above convergence is equivalent to

$$\lim_{n\to\infty} \sum_{k=0}^{n} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \tag{28}$$
$$= E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{\tau^*-1} R(X_t, A_t) + G(X_{\tau^*}) \right) I[\tau^* < \infty] \right].$$

On the other hand, using that R is nonnegative, the monotonicity of U_λ yields that

$$\lim_{n\to\infty} U_\lambda \left(\sum_{k=0}^{n-1} R(X_k, A_k)) \right) = U_\lambda \left(\sum_{k=0}^{\infty} R(X_k, A_k) \right),$$

and combining this convergence with $I[\tau^* > n] \searrow I[\tau^* = \infty]$ as $n \to \infty$, it follows that

$$\lim_{n\to\infty} U_\lambda \left(\sum_{k=0}^{n-1} R(X_k, A_k)) \right) I[\tau^* > n] = U_\lambda \left(\sum_{k=0}^{\infty} R(X_k, A_k)) \right) I[\tau^* = \infty]. \tag{29}$$

Now, assume that $\lambda > 0$. In this context $U_\lambda(\cdot)$ is nonnegative, and via Fatou's lemma it follows that

$$\liminf_{n\to\infty} E_x^\pi \left[U_\lambda \left(\sum_{k=0}^{n-1} R(X_k, A_k)) \right) I[\tau^* > n] \right] \geq E_x^\pi \left[U_\lambda \left(\sum_{k=0}^{\infty} R(X_k, A_k)) \right) I[\tau^* = \infty] \right]. \tag{30}$$

On the other hand, if λ is negative it follows that $\left|U_\lambda\left(\sum_{k=0}^{n-1} R(X_k, A_k)\right)\right| \le 1$, since R is nonnegative, and then

$$\lim_{n\to\infty} E_x^\pi \left[U_\lambda\left(\sum_{k=0}^{n-1} R(X_k, A_k)\right) I[\tau^* > n]\right] = E_x^\pi \left[U_\lambda\left(\sum_{k=0}^{\infty} R(X_k, A_k)\right) I[\tau^* = \infty]\right],$$

by dominated convergence, so that (30) also holds in this case. Now, taking the inferior limit as n goes to ∞ in (27), via (28) and (30) it follows that

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) &\ge E_x^\pi \left[U_\lambda\left(\sum_{t=0}^{\tau^*-1} R(X_t, A_t) + G_\lambda^*(X_{\tau^*})\right) I[\tau^* < \infty]\right] \\ &\qquad + E_x^\pi \left[U_\lambda\left(\sum_{k=0}^{\infty} R(X_k, A_k)\right) I[\tau^* = \infty]\right] \\ &= U_\lambda(V(x; \pi, \tau^*)), \quad x \in S, \quad \pi \in \mathcal{P}; \end{aligned}$$

where (8) was used to set the equality, and the conclusion follows using the strict monotonicity of $U_\lambda(\cdot)$. □

Lemma 4.2. For every $x \in S$ and $\tau \in \mathcal{T}$ assertions (i) and (ii) hold:
(i) For each $n \in \mathbb{N}$,

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) \le E_x^{f^*} &\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + W_\lambda^*(X_\tau)\right) I[\tau \le n]\right] \\ &+ E_x^{f^*} \left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau \ge n+1]\right]. \end{aligned} \tag{31}$$

(ii) $W_\lambda^*(x) \le V_\lambda(x; f^*, \tau)$.

P r o o f . Combining (20) and (21) it follows that

$$U_\lambda(W_\lambda^*(x)) \le \sum_{y\in S} p_{x,y}(f^*(x)) U_\lambda(R(x, f^*(x)) + W_\lambda^*(y)), \quad x \in S, \tag{32}$$

a relation that via the Markov property leads to

$$U_\lambda(W_\lambda^*(X_n)) \le E_x^{f^*} \left[U_\lambda(R(X_n, A_n) + W_\lambda^*(X_{n+1}))\,|\,\mathcal{F}_n\right], \quad x \in S, \quad n \in \mathbb{N}. \tag{33}$$

Next, (31) will be verified by induction. Let $x \in S$ and $\tau \in \mathcal{T}$ be arbitrary. Since τ attains values in $\mathbb{N} \cup \{\infty\}$, combining convention (1) with the equality $P_x^{f^*}[X_0 = x] = 1$

it follows that

$$\begin{aligned}
U_\lambda(W_\lambda^*(x)) &= U_\lambda(W_\lambda^*(X_0))I[\tau=0] + U_\lambda(W_\lambda^*(X_0))I[\tau\geq 1] \\
&= U_\lambda\left(\sum_{t=0}^{0-1} R(X_t,A_t) + W_\lambda^*(X_0)\right) I[\tau=0] + U_\lambda(W_\lambda^*(X_0))I[\tau\geq 1] \\
&\leq U_\lambda\left(\sum_{t=0}^{0-1} R(X_t,A_t) + W_\lambda^*(X_0)\right) I[\tau=0] \\
&\qquad + I[\tau\geq 1]E_x^{f^*}\left[U_\lambda(R(X_0,A_0)+W_\lambda^*(X_1))\,|\,\mathcal{F}_0\right] \\
&= U_\lambda\left(\sum_{t=0}^{0-1} R(X_t,A_t) + W_\lambda^*(X_\tau)\right) I[\tau=0] \\
&\qquad + E_x^{f^*}\left[U_\lambda(R(X_0,A_0)+W_\lambda^*(X_1))I[\tau\geq 1]\,|\,\mathcal{F}_0\right], \quad P_x^{f^*}\text{-a. s.}
\end{aligned}$$

where (33) with $n=0$ was used to set the inequality, and the inclusion $[\tau\geq 1]\in\mathcal{F}_0$ was used in the last step. Taking the expectation with respect to $P_x^{f^*}$, the above display yields the case $n=0$ of (31). Next, assume that $n\in\mathbb{N}$ is such that (31) is valid, and observe that

$$\begin{aligned}
&U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t) + W_\lambda^*(X_{n+1})\right) I[\tau\geq n+1] \\
&\quad = U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t) + W_\lambda^*(X_{n+1})\right) I[\tau=n+1] \\
&\qquad\qquad + U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t) + W_\lambda^*(X_{n+1})\right) I[\tau\geq n+2]
\end{aligned}$$

whereas, using (4),

$$\begin{aligned}
&U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t) + W_\lambda^*(X_{n+1})\right) I[\tau\geq n+2] \\
&\quad = e^{\lambda\sum_{t=0}^{n} R(X_t,A_t)} I[\tau\geq n+2]U_\lambda\left(W_\lambda^*(X_{n+1})\right) \\
&\quad \leq e^{\lambda\sum_{t=0}^{n} R(X_t,A_t)} I[\tau\geq n+2]E_x^{f^*}\left[U_\lambda(R(X_{n+1},A_{n+1})+W_\lambda^*(X_{n+2}))\,|\,\mathcal{F}_{n+1}\right] \\
&\quad = E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n+1} R(X_t,A_t) + W_\lambda^*(X_{n+2})\right) I[\tau\geq n+2]\,\middle|\,\mathcal{F}_{n+1}\right]
\end{aligned}$$

where (33) with $n+1$ instead of n was used to set the inequality, and the second equality was obtained combining (4) with the fact that $e^{\lambda\sum_{t=0}^{n} R(X_t,A_t)}I[\tau\geq n+2]$ is

$\mathcal{F}_{n+1}$-measurable. These two last displays together imply that

$$E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t)+W_\lambda^*(X_{n+1})\right)I[\tau\geq n+1]\right]$$
$$\leq E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t)+W_\lambda^*(X_\tau)\right)I[\tau=n+1]\right]$$
$$+E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n+1} R(X_t,A_t)+W_\lambda^*(X_{n+2})\right)I[\tau\geq n+2]\right],$$

and combining this relation with the induction hypothesis it follows that (31) holds with $n+1$ instead of n, completing the induction argument.

(ii) Consider the following exhaustive cases:

Case 1: $\lambda>0$, so that the increasing function $U_\lambda(\cdot)$ is nonnegative.

In this context, observe that following facts (a) and (b):

(a) Notice that $U_\lambda(\sum_{k=0}^{\infty} R(X_k,A_k))=\infty$, by Assumption 2.1(iv). Thus, if $P_x^{f^*}[\tau=\infty]>0$ then

$$E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\infty} R(X_k,A_k)\right)I[\tau=\infty]\right]=\infty;$$

in this case $\infty=V_\lambda(x;f^*,\tau)$, by (8), and it follows that

$$V_\lambda(x;f^*,\tau)\geq W_\lambda^*(x). \tag{34}$$

(b) If $E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)\right)I[\tau<\infty]\right]=\infty$ then, using that $G(\cdot)\geq 0$,

$$E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+G(X_\tau)\right)I[\tau<\infty]\right]$$
$$\geq E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)\right)I[\tau<\infty]\right]=\infty$$

and via (8) it follows that (34) also holds.

By (a) and (b), to establish (34) in the general case, it is now sufficient to assume that

$$P_x^{f^*}[\tau=\infty]=0 \quad \text{and} \quad E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)\right)I[\tau<\infty]\right]<\infty. \tag{35}$$

Under these conditions, the monotone convergence theorem yields that, as $n\to\infty$,

$$E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W_\lambda^*(X_\tau)\right)I[\tau\leq n]\right] \tag{36}$$
$$\nearrow E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W_\lambda^*(X_\tau)\right)I[\tau<\infty]\right].$$

Next, observe that (4) and (35) together imply that

$$\begin{aligned} E_x^{f^*}&\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W_\lambda^*(X_\tau)\right)I[\tau<\infty]\right] \\ &\leq E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+\|W_\lambda^*\|\right)I[\tau<\infty]\right] \\ &\leq e^{\lambda\|W_\lambda^*\|}E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)\right)I[\tau<\infty]\right]<\infty, \end{aligned}$$

and then (36) immediately yields that

$$\begin{aligned} E_x^{f^*}&\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W_\lambda(X_\tau)\right)I[\tau\geq n+1]\right] \\ &= E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W_\lambda(X_\tau)\right)I[\tau<\infty]\right] \\ &\qquad - E_x^{f^*}\left[U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W_\lambda(X_\tau)\right)I[\tau\leq n]\right] \\ &\to 0 \text{ as } n\to\infty. \end{aligned} \tag{37}$$

On the other hand, using (4) repeatedly, observe that

$$\begin{aligned} 0\leq E_x^{f^*}&\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t)+W_\lambda^*(X_{n+1})\right)I[\tau\geq n+1]\right] \\ &\leq E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t)+\|W_\lambda^*\|\right)I[\tau\geq n+1]\right] \\ &\leq e^{\lambda\|W_\lambda^*\|}E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t)\right)I[\tau\geq n+1]\right] \\ &\leq e^{\lambda\|W_\lambda^*\|}E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t)\right)I[\tau\geq n+1]\right] \\ &\leq e^{\lambda\|W_\lambda^*\|}E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t)+G(X_\tau)\right)I[\tau\geq n+1]\right], \end{aligned}$$

where the third inequality is due to the fact that $U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t)\right)\geq U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t)\right)$ on the event $[\tau\geq n+1]$, and the fourth inequality is due to the nonnegativity of G. Thus,

$$E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t)+W_\lambda^*(X_{n+1})\right)I[\tau\geq n+1]\right]\to 0 \text{ as } n\to\infty.$$

Taking the limit as n goes to ∞ in both sides of (31), the above convergence and (36) together imply that

$$W_\lambda^*(x) \le E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + W_\lambda^*(X_\tau)\right) I[\tau<\infty]\right]$$
$$\le E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + G(X_\tau)\right) I[\tau<\infty]\right]$$

where the inequality $W_\lambda^*(\cdot) \le G(\cdot)$ was used in the last step. Recalling that the condition that $P_x^{f^*}[\tau=\infty]=0$ is in force, the above relation and (8) together yield that $W_\lambda^*(x) \le V_\lambda(x;f^*,\tau)$.

Case 2: $\lambda < 0$, so that $|U_\lambda(\cdot)|$ is decreasing and $U_\lambda(x) \in [-1,0)$ when $x \in [0,\infty)$, by (3).

In this context, Assumption 2.1(iv) yields that

$$\left|U_\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k) + W(X_\tau)\right)\right| = e^{\lambda\left(\sum_{k=0}^{\tau-1} R(X_k,A_k)+W(X_\tau)\right)}$$
$$\le e^{\lambda\|W\|}\left(e^{\lambda\delta}\right)^{n+1} \text{ on the event } [\tau \ge n+1],$$

and it follows that

$$E_x^{f^*}\left[\left|U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t) + W_\lambda^*(X_{n+1})\right)\right| I[\tau\ge n+1]\right] \le e^{\lambda\|W_\lambda^*\|}\left(e^{\lambda\delta}\right)^{n+1} \to 0 \text{ as } n\to\infty.$$

On the other hand, observe that $\left|U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + W_\lambda^*(X_\tau)\right)\right| \le 1$, since R and W_λ^* are nonnegative, and the the dominated convergence theorem implies that

$$\lim_{n\to\infty} E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + W_\lambda^*(X_\tau)\right) I[\tau\le n]\right]$$
$$= E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + W_\lambda^*(X_\tau)\right) I[\tau<\infty]\right]$$

Taking the limit as n goes to ∞ in (31) the two last displays together imply that

$$U_\lambda(W_\lambda^*(x)) \le E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + W_\lambda^*(X_\tau)\right) I[\tau<\infty]\right]$$
$$\le E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t,A_t) + G(X_\tau)\right) I[\tau<\infty]\right]$$

where the second inequality is due to the inclusion $W_\lambda^* \in [\![0,G]\!]$. Finally, observe that Assumption 2.1(iv) implies that $U_\lambda\left(\sum_{t=0}^{\infty} R(X_t,A_t)\right)=0$, and combining this equality with the above display and (8) it follows that $U_\lambda(W_\lambda^*(x)) \le U_\lambda(V_\lambda(x,f^*,\tau))$, so that $W_\lambda^*(x) \le V_\lambda(x,f^*,\tau)$. □

5. PROOF OF THE MAIN RESULT

After the preliminaries in the previous section, Theorem 3.3 can be established as follows.

Proof of the Theorem 3.3 Notice that parts (i) and (ii) follows from Lemmas 4.1 and 4.2, respectively. Next, parts (i) and (ii) together lead to

$$V_\lambda(\cdot;\pi,\tau^*) \leq W_\lambda^*(\cdot) \leq V_\lambda(\cdot;f^*,\tau), \quad (\pi,\tau) \in \mathcal{P}\times\mathcal{T},$$

and setting $(\pi,\tau) = (f^*,\tau^*)$ it follows that $W_\lambda^*(\cdot) = V_\lambda(\cdot;f^*,\tau^*)$, so that (f^*,τ^*) is a Nash equilibrium, by Definition 2.2. □

6. CONCLUSION

In this work, Markov stopping games on a denumerable state space were studied. The performance of a pair of strategies was measured by the risk-sensitive total reward criterion, and the problem of establishing the existence of a Nash equilibrium was analyzed. Apart from standard continuity-compactness requirements, the basic conditions on the system concern with the reward structure: The terminal reward G is nonnegative, and the running reward R attains values in a compact interval contained in $(0,\infty)$. The main result of the paper, stated in Theorem 3.3, establishes that the value of the game is the fixed point W_λ^* of the operator T_λ in Definition 3.1, and that W_λ^* determines a Nash equilibrium. Extending the results in this work to more general frameworks seems to be an interesting problem.

ACKNOWLEDGEMENT

The authors are deeply grateful to the reviewers and the Associate Editor for their careful reading of the original manuscript and for their advice to improve the paper. This work was partially supported by the PSF Organization under Grant No. 030711. Manuel A. Torres–Gomar and Hugo Cruz–Suárez dedicate this article to the memory of their collaborator and co-author of the present work, Rolando Cavazos–Cadena, whose unfortunate death occurred on May 24, 2023.

(Received October 22, 2023)

REFERENCES

[1] A. Alanís-Durán and R. Cavazos-Cadena: An optimality system for finite average Markov decision chains under risk-aversion. Kybernetika *48* (2012), 1, 83–104.

[2] N. Bäuerle and U. Rieder: Markov Decision Processes with Applications to Finance. Springer–Verlag, New York 2011.

[3] N. Bäuerle and U. Rieder: More risk-sensitive Markov decision processes. Math. Oper. Res. *39* (2014), 1, 105–120. DOI:10.1287/moor.2013.0601

[4] S. Balaji and S. P. Meyn: Multiplicative ergodicity and large deviations for an irreducible Markov chain. Stoch. Proc. Appl. *90* (2000), 1, 123–144. DOI:10.1016/S0304-4149(00)00032-6

[5] 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

[6] V. S. Borkar and S. P. Meyn: Risk-sensitive optimal control for Markov decision process with monotone cost. Math. Oper. Res. *27* (2002), 1, 192-209. DOI:10.1287/moor.27.1.192.334

[7] R. Cavazos-Cadena and D. Hernández–Hernández: A system of Poisson equations for a non-constant Varadhan functional on a finite state space. Appl. Math. Optim. *53* (2006), 101–119. DOI:10.1007/s00245-005-0840-3

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

[9] R. Cavazos-Cadena, L. Rodríguez–Gutiérrez and D. M. Sánchez–Guillermo: Markov stopping game with an absorbing state. Kybernetika *57* (2021), 3, 474–492. DOI:10.14736/kyb-2021-3-0474

[10] R. Cavazos-Cadena, M. Cantú–Sifuentes and I. Cerda–Delgado: Nash equilibria in a class of Markov stopping games with total reward criterion. Math. Methods Oper. Res. *94* (2021), 319–340. DOI:10.1007/s00186-021-00759-5

[11] E. V. Denardo and U. G. Rothblum: A turnpike theorem for A risk-sensitive Markov decision process with stopping. SIAM J. Control Optim. *45* (2006), 2, 414–431. DOI:10.1137/S0363012904442616

[12] G. B. Di Masi and L. Stettner: Risk-sensitive control of discrete-time Markov processes with infinite horizon. SIAM J. Control Optim. *38* (1999), 1, 61–78. DOI:10.1137/S0363012997320614

[13] G. B. Di Masi and L. Stettner: Infinite horizon risk sensitive control of discrete time Markov processes with small risk. Systems Control Lett. *40* (2000), 1, 305–321. DOI:10.1016/S0167-6911(00)00018-9

[14] G. B. Di Masi and L. Stettner: Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. SIAM J. Control Optim. *46* (2007), 1, 231–252. DOI:10.1137/040618631

[15] O. Hernández-Lerma: Adaptive Markov Control Processes. Springer, New York 1988.

[16] R. Howard and J. Matheson: Risk-sensitive Markov decision processes. Management Science *18* (1972), 356–369. DOI:10.1287/mnsc.18.7.356

[17] A. Jaśkiewicz: Average optimality for risk sensitive control with general state space. Ann. App. Probab. *17* (2007), 2, 654–675. DOI:10.1214/105051606000000790

[18] I. Kontoyiannis and S. P. Meyn: Spectral theory and limit theorems for geometrically ergodic Markov processes. Ann. App. Probab. *13* (2003), 1, 304–362. DOI:10.1214/aoap/1042765670

[19] J. López-Rivero, R. Cavazos-Cadena, and H. Cruz-Suárez: Risk-sensitive Markov stopping games with an absorbing state. Kybernetika *58* (2022), 1, 101–122. DOI:10.14736/kyb-2022-1-0101

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

[21] M. Pitera and L. Stettner: Long run risk sensitive portfolio with general factors. Math. Meth. Oper. Res. *82* (2016), 2, 265–293. DOI:10.1007/s00186-015-0514-0

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

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

[24] K. Sladký: Risk-sensitive Ramsey growth model. In: Proce. 27th International Conference Mathematical Methods in Economics 2010 (M. Houda and J. Friebelová, eds.), Ceské Budějovice 2010, pp. 1–6.

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

[26] L. Stettner: Risk sensitive portfolio optimization. Math. Meth. Oper. Res. *50* (1999), 3, 463–474. DOI:10.1007/s001860050081

Manuel A. Torres-Gomar, Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Ave. San Claudio y Río Verde, Col. San Manuel CU, Puebla, PUE, 72570. México.

e-mail: manuel.torresg@alumno.buap.mx

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

Hugo Cruz-Suárez, Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Ave. San Claudio y Río Verde, Col. San Manuel CU, Puebla, PUE, 72570. México.

e-mail: hcs@fcfm.buap.mx