



Contractive approximations in average Markov decision chains driven by a risk-seeking controller

Gustavo Portillo-Ramírez¹ · Rolando Cavazos-Cadena² · Hugo Cruz-Suárez¹ 

Received: 8 November 2022 / Revised: 24 June 2023 / Accepted: 28 June 2023 /

Published online: 5 July 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

This work concerns with Markov decision processes on a denumerable state space. It is assumed that the performance of a control policy is measured by the average criterion associated with a risk-seeking controller with constant risk-sensitivity coefficient. The structural assumptions on the model ensure that the optimal average cost is constant, but it is possible that the optimality equation does not admit a solution. In this context, a risk-sensitive version of the classical *discounted approach* is used to obtain convergent approximations to the optimal average cost, and to determine *nearly optimal* stationary policies.

Keywords Risk-lover decision maker · Exponential utility · Contractive operator · Fixed point · Hölder's inequality

Mathematics Subject Classification 93E20 · 93C55 · 60J05

1 Introduction

This note concerns with Markov decision chains evolving on a denumerable state space. The one-step cost function is bounded and the performance of a control policy is measured by the average criterion associated with a risk-seeking decision maker. The structural conditions on the transition law ensure that the optimal average cost is

✉ Hugo Cruz-Suárez
hcs@fcfm.buap.mx

Gustavo Portillo-Ramírez
gustavo.portillo@alumno.buap.mx

Rolando Cavazos-Cadena
rolando.cavazos@uaaan.edu.mx

¹ Benemérita Universidad Autónoma de Puebla, Puebla, PUE, Mexico

² Universidad Autónoma Agraria Antonio Narro, Saltillo, COAH, Mexico

constant, but do not guarantee that the optimality equation admits a solution. In this framework, the following problem is addressed:

- To obtain convergent approximations to the optimal average cost, and to determine approximately optimal stationary policies using the fixed points of a family of contractive operators.

The main conclusions on this problem, which are stated in Theorem 3.1 of Sect. 3, represent an extension of the classical ‘discounted approach’ in the risk-neutral case (Hernández-Lerma 1989; Arapostathis et al. 1993), and extend to the present framework results established in Saucedo-Zul et al. (2020), where a risk-averse version of this problem was analyzed.

The study of Markov decision chains endowed with a risk-sensitive average criterion can be traced back, at least, to the seminal paper by Howard and Matheson (1972), where Markov decision chains with finite state space were analyzed, and the optimal average cost was characterized via an optimality equation. The interest on this topic has been motivated by applications, for instance, in finance (Bäuerle and Rieder 2011, 2014; Stettner 1999; Pitera and Stettner 2016), revenue management (Barz and Waldmann 2007), and the theory of large deviations (Borkar and Meyn 2002). Models with finite or denumerable state space are considered, for instance, in Sladký (2008, 2018), Cavazos-Cadena (2009, 2018) whereas Markov decision chains on a Borel states space are analyzed in Di Masi and Stettner (1999, 2000, 2007), Jaśkiewicz (2007), Jaśkiewicz and Nowak (2014) and Shen et al. (2013).

Stochastic games with risk-sensitive criteria are studied in Basu and Ghosh (2014).

The remainder of the paper is organized as follows. In Sect. 2 the decision model is formally described, the average criterion is defined, and the main structural assumptions on the model are stated. In Sect. 3 a family of contractive operators is introduced, and the main result of the paper is stated as Theorem 3.1. The technical instruments that will be used to establish that result are established in Sect. 4, and the proof of the main result is presented in Sect. 5 before the concluding remarks.

Notation Throughout the remainder \mathbb{N} denotes the set of non-negative integers and, given a topological space S , the Banach space of all bounded functions $H : S \rightarrow \mathbb{R}$ is denoted by $\mathcal{B}(S)$; the supremum norm of $H \in \mathcal{B}(S)$ is denoted by $\|H\| := \sup_{x \in S} |H(x)|$. On the other hand, every (in)equality involving random variables holds almost surely with respect to the underlying probability measure.

2 Decision model

Let $\mathcal{M} := (S, A, \{A(x)\}_{x \in S}, C, [p_{x,y}(a)])$ be a Markov decision chain, a model for a dynamical system whose components are as follows: The state space S is a denumerable set endowed with the discrete topology, the metric space A is the action set whereas, for each state $x \in S$, $A(x) \subset A$ is the class of admissible actions (controls) at state x . On the other hand $C : \mathbb{K} \rightarrow \mathbb{R}$ is the cost function, where $\mathbb{K} = \{(x, a) \mid x \in S, a \in A(x)\}$ is the family of admissible pairs and, finally, $[p_{x,y}(a)]_{x,y \in S, a \in A(x)}$ is the controlled transition law. The interpretation of \mathcal{M} is as follows: At each time $t \in \mathbb{N}$ the decision maker observes the state of the system $X_t = x \in S$, and then picks

and applies an action $A_t = a \in A(x)$. As a consequence of such an intervention, (i) a cost $C(x, a)$ is incurred, and (ii) the system moves to a new state $X_{t+1} \in S$ where, regardless of the previous states and actions, the event $[X_{t+1} = y]$ is observed with probability $p_{x,y}(a)$, where $\sum_{y \in S} p_{x,y}(a) = 1$; this is the Markov property of the decision process.

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

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

(iii) The cost function is bounded, i.e., $C \in \mathcal{B}(\mathbb{K})$.

Policies A control policy is a rule for choosing actions, which at each decision time $n \in \mathbb{N}$ may depend on the current state as well as the previous states and actions. More formally, for each $n \in \mathbb{N}$ define the space \mathbb{H}_n of possible histories up to time n by $\mathbb{H}_0 := S$ and $\mathbb{H}_n := \mathbb{K}^n \times S$ for $n = 1, 2, 3, \dots$; a generic elements of \mathbb{H}_n is denoted by $h_n = (x_0, a_0, x_1, a_1, \dots, x_{n-1}, a_{n-1}, x_n)$, where $(x_k, a_k) \in \mathbb{K}$ for $k < n$ and $x_n \in S$. With this notation, a control policy $\pi = \{\pi_n\}$ is a sequence of stochastic kernels π_n on A given \mathbb{H}_n , satisfying that $\pi_n(A(x_n)|h_n) = 1$, for each $h_n \in \mathbb{H}_n$ and $n \in \mathbb{N}$. The family of all policies is denoted by \mathcal{P} . Next, set $\mathbb{F} := \prod_{x \in S} A(x)$, which is compact metric space, by Assumption 2.1, and consists of all functions $f : S \rightarrow A$ satisfying $f(x) \in A(x)$ for every $x \in S$. A policy $\pi \in \mathcal{P}$ is stationary if there exists $f \in \mathbb{F}$ such that the equality $\pi_n(\{f(x_n)\}|h_n) = 1$ always holds: the class of stationary policies is naturally identified with \mathbb{F} , a convention allowing to write $\mathbb{F} \subset \mathcal{P}$. Given the initial state $X_0 = x$ and the policy $\pi \in \mathcal{P}$ used to drive the system, the distribution of the state-action process $\{(X_t, A_t)\}_{t \in \mathbb{N}}$ is uniquely determined and is denoted by P_x^π (Hernández-Lerma 1989; Arapostathis et al. 1993; Puterman 1994), whereas E_x^π stands for the corresponding expectation operator.

Throughout the sequel, the following notation will be used: For each $n \in \mathbb{N}$ set

$$H_n := (X_0, A_0, \dots, X_{n-1}, A_{n-1}, X_n) \text{ and } \mathcal{F}_n := \sigma(H_n), \tag{2.1}$$

whereas for each $F \subset S$ the first return time to set F is defined by

$$T_F := \min\{n \geq 1 \mid X_n \in F\}; \tag{2.2}$$

when $F = \{x\}$ is a singleton the simpler notation

$$T_x \equiv T_{\{x\}} \tag{2.3}$$

is used. Notice that T_F is an stopping time respect to the filtration $\{\mathcal{F}_n\}$, i. e., $[T_F = n] \in \mathcal{F}_n$ for every $n \in \mathbb{N}$.

Average criterion Throughout the remainder it is supposed that the decision maker has a constant risk-sensitive coefficient λ which satisfies

$$\lambda < 0.$$

This means that the controller assesses a random cost Y via the expectation of $U_\lambda(Y)$, where the (dis-)utility function $U_\lambda : \mathbb{R} \rightarrow (-\infty, 0)$ is defined as follows

$$U_\lambda(x) = -e^{\lambda x}, \quad x \in \mathbb{R}; \quad (2.4)$$

notice that $U_\lambda(\cdot)$ is strictly increasing and satisfies the relation

$$U_\lambda(a + b) = e^{\lambda a} U_\lambda(b), \quad a, b \in \mathbb{R}. \quad (2.5)$$

When the decision maker chooses between two random costs C_0 and C_1 , the controller prefers C_0 if $E[U_\lambda(C_0)] < E[U_\lambda(C_1)]$, and is indifferent between both costs if $E[U_\lambda(C_0)] = E[U_\lambda(C_1)]$. The certainty equivalent of a cost Y is denoted by $\mathcal{E}_\lambda[Y]$ and is determined by the equality $U_\lambda(\mathcal{E}_\lambda[Y]) = E[U_\lambda(Y)]$, so that the controller is indifferent between paying the fixed amount $\mathcal{E}_\lambda(Y)$ or facing the random cost Y . Notice that $U_\lambda(\cdot)$ is a concave function, so that Jensen's inequality yields that $\mathcal{E}_\lambda(Y) \leq E[Y]$. Now, observe that

$$\mathcal{E}_\lambda[Y] = U_\lambda^{-1}(E[U_\lambda(Y)]) = \frac{1}{\lambda} \log \left(E \left[e^{\lambda Y} \right] \right), \quad (2.6)$$

an expression that immediately yields that

$$P[|Y| \leq b] = 1 \implies |\mathcal{E}_\lambda(Y)| \leq b. \quad (2.7)$$

Next, assume that the controller chooses actions using policy $\pi \in \mathcal{P}$ starting at $x \in S$. The application of the first n actions A_0, A_1, \dots, A_{n-1} generates the cost $\sum_{k=0}^{n-1} C(X_k, A_k)$ and, by (2.6), the associated certainty equivalent is given by

$$J_n(\pi, x) := \frac{1}{\lambda} \log \left(E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right), \quad n = 1, 2, 3, \dots, \quad (2.8)$$

which represents an average of $J_n(\pi, x)/n$ per step. The (inferior limit λ -sensitive) average performance index of policy $\pi \in \mathcal{P}$ at state $x \in S$ is given by

$$J(\pi, x) := \liminf_{n \rightarrow \infty} \frac{1}{n} J_n(\pi, x), \quad (2.9)$$

and

$$J_*(x) := \inf_{\pi \in \mathcal{P}} J(\pi, x), \quad x \in S. \quad (2.10)$$

is the corresponding optimal value function. A policy $\pi_* \in \mathcal{P}$ is (λ -)average optimal if $J(\pi, x) = J(\pi_*, x)$ for every $x \in S$.

Recurrence-communication conditions In the risk-neutral case, it is known that the *simultaneous Doeblin condition*, which is stated in Assumption 2.2(i) below, is sufficient to ensure that the optimal average cost is constant and is characterized via

an optimality equation (Hernández-Lerma 1989; Arapostathis et al. 1993; Puterman 1994). In the present risk-sensitive context, the λ -sensitive average optimality equation is given by

$$U_\lambda(g + h(x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{x,y}(a) U_\lambda(C(x, a) + h(y)) \right], \quad x \in S, \quad (2.11)$$

where g is a real number and $h : S \rightarrow \mathbb{R}$ is a function. When this equation admits a solution $(g, h(\cdot))$ and $h(\cdot)$ is a bounded mapping, it is known that the optimal λ -average cost function $J_*(\cdot)$ is constant and equal to g , and if $f \in \mathbb{F}$ is such that for each state x action $f(x)$ minimizes the term within brackets in (2.11), then f is λ -average optimal; see, for instance, Howard and Matheson (1972), Hernández-Hernández and Marcus (1996), or Cavazos-Cadena (2009). Notice that via (2.4) the above optimality equation can be equivalently written as

$$e^{\lambda g + \lambda h(x)} = \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \right], \quad x \in S. \quad (2.12)$$

In contrast with the risk-neutral context, in the present framework where the controller is risk-seeking, the simultaneous Doeblin conditions is not sufficient to ensure even that the optimal average cost function is constant (Cavazos-Cadena and Fernández-Gaucherand 1999; Cavazos-Cadena 2009). For this reason, in this work the simultaneous Doeblin condition will be complemented with a communication requirement.

Assumption 2.2 There exists $z \in S$ such that properties (i) and (ii) below hold:

(i) [Simultaneous Doeblin Condition.] The first return time T_z satisfies

$$\sup_{x \in S, f \in \mathbb{F}} E_x^f [T_z] < \infty. \quad (2.13)$$

(ii) [Accessibility from z .] Under the action of any stationary policy, every state $y \in S$ is accessible from z , that is,

$$P_z^f [T_y < \infty] > 0, \quad y \in S, \quad f \in \mathbb{F}. \quad (2.14)$$

Remark 2.1 Assumptions 2.1 and 2.2 imply the following properties (i) and (ii) below; for a proof see Theorem 4.1 in Cavazos-Cadena (2018).

(i) For each $y \in S$, there exists a finite constant M_y such that

$$E_x^\pi [T_y] \leq M_y, \quad x \in S, \quad \pi \in \mathcal{P}. \quad (2.15)$$

(ii) If $x, y \in S$ with $x \neq y$, then $P_x^\pi [T_y < T_x] > 0$ for every $\pi \in \mathcal{P}$.

Remark 2.2 Assumption 2.2 is, admittedly, very strong. However, in the denumerable case such a condition is presently the most general one under which a characterization of the optimal risk-sensitive average cost is available. The result in this direction can be seen in Cavazos-Cadena (2018) and involves an extension of the Collatz-Wielandt relations in the theory of positive matrices.

The Problem Under Assumptions 2.1 and 2.2 the optimal average cost function $J_*(\cdot)$ is constant but the optimality equation (2.11) does not necessarily admits a solution; an (uncontrolled) example illustrating this phenomenon was presented in Section 9 of Cavazos-Cadena (2018). This fact provides that motivation to analyze the following problem:

- To obtain convergent approximations to the optimal average cost as well as ‘nearly optimal’ stationary policies via the fixed points of contractive operators.

An answer to this problem allows to determine approximations to the optimal average cost as well as a stationary policy whose average cost is ‘close’ to the optimal one by solving the single equation characterizing the fixed point of a contractive operator. The main result on the above problem is stated in the following section, and represents an extension of the classical ‘discounted approach’ in the risk neutral case (Hernández-Lerma 1989; Puterman 1994) to the present risk-seeking framework.

Throughout the remainder, even without explicit reference, Assumptions 2.1 and 2.2 are enforced.

3 Contractive approximations

In this section the main result of the paper will be stated in Theorem 3.1 below. To begin with, for each $\alpha \in (0, 1)$ define $T_\alpha: \mathcal{B}(S) \rightarrow \mathcal{B}(S)$ as follows: For each $W \in \mathcal{B}(S)$, $T_\alpha[W]$ is implicitly determined by

$$U_\lambda(T_\alpha[W](x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{x,y}(a) U_\lambda(C(x, a) + \alpha W(y)) \right], \quad x \in S, \quad (3.1)$$

an expression that via (2.4) leads to

$$T_\alpha[W](x) := \frac{1}{\lambda} \log \left(\sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda \alpha W(y)} \right] \right), \quad x \in S. \quad (3.2)$$

Using (2.7) it follows that $\|T_\alpha[W]\| \leq \|C\| + \alpha \|W\|$, so that T_α maps $\mathcal{B}(S)$ into itself. Also, it is not difficult to verify that T_α is a monotone and α -homogeneous operator, i. e., for each $W, V \in \mathcal{B}(S)$

$$W \geq V \implies T_\alpha[W] \geq T_\alpha[V] \text{ and } T_\alpha[V + c] = T_\alpha[V] + \alpha c, \quad c \in \mathbb{R}. \quad (3.3)$$

Observing that $V \leq W + \|V - W\|$, these properties lead to $T_\alpha[V] \leq T_\alpha[W + \|V - W\|] = T_\alpha[W] + \alpha\|V - W\|$, and interchanging the roles of V and W it follows that

$$\|T_\alpha[W] - T_\alpha[V]\| \leq \alpha\|W - V\|, \quad W, V \in \mathcal{B}(S), \tag{3.4}$$

so that T_α is a contractive operator on $\mathcal{B}(S)$. Since $\mathcal{B}(S)$ endowed with the supremum norm is a Banach space, there exists a unique $V_\alpha \in \mathcal{B}(S)$ satisfying

$$V_\alpha = T_\alpha[V_\alpha], \tag{3.5}$$

an equation that, via (3.2), is equivalent to

$$e^{\lambda V_\alpha(x)} = \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda \alpha V_\alpha(y)} \right], \quad x \in S. \tag{3.6}$$

Additionally, from Assumption 2.1 it is not difficult to see that there exists $f_\alpha \in \mathbb{F}$ such that, for every $x \in S$, action $f_\alpha(x)$ maximizes the term within brackets in the above display, so that

$$e^{\lambda V_\alpha(x)} = e^{\lambda C(x, f_\alpha(x))} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda \alpha V_\alpha(y)}, \quad x \in S. \tag{3.7}$$

The normalized (α)-cost and the (α)-relative value functions are defined by

$$g_\alpha(x) := (1 - \alpha)V_\alpha(x), \quad h_\alpha(x) := \alpha[V_\alpha(x) - V_\alpha(w)], \quad x \in S, \tag{3.8}$$

respectively, where, from this point onwards, $w \in S$ is an arbitrary but fixed state. Direct calculations combining these definitions with the two previous displays yield that

$$e^{\lambda g_\alpha(x) + \lambda h_\alpha(x)} = \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_\alpha(y)} \right], \quad x \in S, \tag{3.9}$$

and

$$e^{\lambda g_\alpha(x) + \lambda h_\alpha(x)} = e^{\lambda C(x, f_\alpha(x))} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda h_\alpha(y)}, \quad x \in S. \tag{3.10}$$

Notice that $\|V_\alpha - T_\alpha[0]\| = \|T_\alpha[V_\alpha] - T_\alpha[0]\| \leq \alpha\|V_\alpha - 0\| = \alpha\|V_\alpha\|$, and then, observing that $\|T_\alpha[0]\| \leq \|C\|$, by (3.2), it follows that $\|V_\alpha\| - \|C\| \leq \|V_\alpha\| - \|T_\alpha[0]\| \leq \|V_\alpha - T_\alpha[0]\| \leq \alpha\|V_\alpha\|$, so that

$$\|g_\alpha\| = (1 - \alpha)\|V_\alpha\| \leq \|C\|. \tag{3.11}$$

The next theorem is the main result of this work.

Theorem 3.1 *Let $\lambda < 0$ be arbitrary but fixed. Under Assumptions 2.1 and 2.2 the following assertions (i) and (ii) hold.*

- (i) *The optimal average cost is constant, say g^* , and $\lim_{\alpha \nearrow 1} g_\alpha(x) = g^* = J_*(x)$ for every $x \in S$.*
 (iii) *Given $\varepsilon > 0$, for each $x \in S$ there exists $\alpha_{x,\varepsilon} \in (0, 1)$ such that policy f_α in (3.7) is ε -optimal at x for $\alpha \in (\alpha_{x,\varepsilon}, 1)$, that is,*

$$\alpha \in (\alpha_{x,\varepsilon}, 1) \implies g^* + \varepsilon \geq J(f_\alpha, x). \quad (3.12)$$

The proof of Theorem 3.1 will be presented in Sect. 5 after the preliminary results established in the following section.

4 Auxiliary tools

In this section the basic technical instruments that will be used to verify Theorem 3.1 are analyzed. Such preliminaries are established in Lemmas 4.1–4.3 below. The first one concerns with boundedness properties of the family of relative cost functions introduced in (3.8).

Lemma 4.1 (i) *For each $\alpha \in (0, 1)$,*

$$h_\alpha(\cdot) \leq 2\|C\|M_w, \quad (4.1)$$

where the finite constant M_w is as in (2.15).

(ii) *For each $x \in S$, $\liminf_{\alpha \nearrow 1} h_\alpha(x) > -\infty$.*

Proof (i) Given $\alpha \in (0, 1)$, define the sequence $\{Y_n\}$ of random variables by $Y_0 = e^{\lambda h_\alpha(X_0)}$ and $Y_n = e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_n)}$ for $n \geq 1$. Now, let $x \in S$ be a fixed state, and observe that (3.10) implies that for every $n \in \mathbb{N}$

$$\begin{aligned} e^{\lambda h_\alpha(X_n)} &= e^{\lambda (C(X_n, f_\alpha(X_n)) - g_\alpha(X_n))} \sum_{y \in S} p_{X_n, y}(f_\alpha(X_n)) e^{\lambda h_\alpha(y)} \\ &= E_x^{f_\alpha} \left[e^{\lambda (C(X_n, A_n) - g_\alpha(X_n)) + \lambda h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right], \quad P_x^{f_\alpha} \text{-a. s.}, \end{aligned} \quad (4.2)$$

where, using that the relation $P_x^{f_\alpha}[A_t = f_\alpha(X_t)] = 1$ is always valid, the second equality is due to the Markov property. Observing that $e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t))}$ is \mathcal{F}_n -measurable, by (2.1), the previous display yields

$$\begin{aligned} Y_n &= e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_n)} \\ &= e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t))} E_x^{f_\alpha} \left[e^{\lambda (C(X_n, A_n) - g_\alpha(X_n)) + \lambda h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \\ &= E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] = E_x^{f_\alpha} [Y_{n+1} | \mathcal{F}_n], \end{aligned}$$

so that $\{(Y_n, \mathcal{F}_n)\}$ is a martingale with respect to $P_x^{f_\alpha}$; since $P_x^{f_\alpha}[X_0 = x] = 1$, the optional sampling theorem yields that, for every initial state x and $n \in \mathbb{N}$,

$$\begin{aligned} e^{\lambda h_\alpha(x)} &= E_x^{f_\alpha}[Y_0] \\ &= E_x^{f_\alpha}[Y_{n \wedge T_w}] = E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + h_\alpha(X_{n \wedge T_w})} \right]. \end{aligned}$$

Now, using (2.2) and (2.3), observe that $h_\alpha(X_{T_w}) = h_\alpha(w) = 0$ on the event $[T_w < \infty]$; since $P_x^{f_\alpha}[T_w < \infty] = 1$, by (2.15), it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} e^{\lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + h_\alpha(X_{n \wedge T_w})} \\ = e^{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + h_\alpha(X_{T_w})} \\ = e^{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t))}, \quad P_x^{f_\alpha}\text{-a. s.} \end{aligned}$$

Via Fatou’s lemma and Jensen’s inequality, these two last displays together imply that

$$\begin{aligned} e^{\lambda h_\alpha(x)} &= \liminf_{n \rightarrow \infty} E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + h_\alpha(X_{n \wedge T_w})} \right] \\ &\geq E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t))} \right] \geq e^{E_x^{f_\alpha} \left[\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) \right]} \\ &\geq e^{E_x^{f_\alpha} \left[-\sum_{t=0}^{T_w - 1} |\lambda(C(X_t, A_t) - g_\alpha(X_t))| \right]} \geq e^{2\lambda \|C\| E_x^{f_\alpha}[T_w]}, \end{aligned}$$

where (3.11) and the negativity of λ were used in the last step. It follows that $\lambda h_\alpha(x) \geq 2\lambda \|C\| E_x^{f_\alpha}[T_w]$, so that $h_\alpha(x) \leq 2\|C\| E_x^{f_\alpha}[T_w]$; since x was arbitrary in this argument, (4.1) follows via (2.15).

(ii) Let $\tilde{f} \in \mathbb{F}$ be fixed, and define the sequence $\{S_k\}$ of subsets of the state space S by

$$\begin{aligned} S_0 &:= \{w\}, \\ S_k &:= \{y \in S : p_{x,y}(\tilde{f}(x)) > 0 \text{ for some } x \in S_{k-1}\}, \quad k = 1, 2, 3, \dots \end{aligned}$$

and notice that $\bigcup_{k=0}^\infty S_k = S$, by Remark 2.1(ii). Thus, to establish part (ii) it is sufficient to show that, for every $k \in \mathbb{N}$,

$$\liminf_{\alpha \nearrow 1} h_\alpha(x) > -\infty, \quad x \in S_k, \tag{4.3}$$

a claim will be verified by induction. To begin with, let $\tilde{f} \in \mathbb{F}$ be a fixed policy and notice that (3.9) implies that

$$\begin{aligned}
 e^{\lambda h_\alpha(x)} &\geq e^{\lambda C(x, \tilde{f}(x)) - g_\alpha(x)} \sum_{y \in S} p_{x,y}(\tilde{f}(x)) e^{\lambda h_\alpha(y)} \\
 &\geq e^{2\lambda \|C\|} \sum_{y \in S} p_{x,y}(\tilde{f}(x)) e^{\lambda h_\alpha(y)}
 \end{aligned}
 \tag{4.4}$$

where the second inequality is due to (3.11) and the negativity of λ . Now, using that $S_0 = \{w\}$ and $h_\alpha(w) = 0$ for every $\alpha \in (0, 1)$, observe that assertion (4.3) clearly holds for $k = 0$. Next, assume that (4.3) is valid for some $k \in \mathbb{N}$ and let $\tilde{y} \in S_{k+1}$ be arbitrary. Pick $\tilde{x} \in S_k$ such that

$$p_{\tilde{x}, \tilde{y}}(\tilde{f}(\tilde{x})) > 0$$

and notice that (4.4) implies that $e^{\lambda h_\alpha(\tilde{x})} \geq e^{2\lambda \|C\|} p_{\tilde{x}, \tilde{y}}(\tilde{f}(\tilde{x})) e^{\lambda h_\alpha(\tilde{y})}$, so that

$$h_\alpha(\tilde{x}) \leq 2\|C\| + \frac{1}{\lambda} \log(p_{\tilde{x}, \tilde{y}}(\tilde{f}(\tilde{x}))) + h_\alpha(\tilde{y}).$$

Since $\tilde{x} \in S_k$, the induction hypothesis yields that $\liminf_{\alpha \nearrow 1} h_\alpha(\tilde{x}) > -\infty$, and then the two last displays together imply that $\liminf_{\alpha \nearrow 1} h_\alpha(\tilde{y}) > -\infty$. Recalling that $\tilde{y} \in S_{k+1}$ is arbitrary, it follows that (4.3) holds with $k + 1$ instead of k , completing the induction argument. \square

In the subsequent development $\{\alpha_n\} \subset (0, 1)$ is a fixed sequence such that

$$\alpha_n \nearrow 1 \text{ as } n \rightarrow \infty \tag{4.5}$$

and, after taking a subsequence—if necessary—without loss of generality it is assumed that the following limits exist:

$$g(x) := \lim_{n \rightarrow \infty} g_{\alpha_n}(x), \quad h^*(x) := \lim_{n \rightarrow \infty} h_{\alpha_n}(x), \quad x \in S \tag{4.6}$$

where, for each $x \in S$,

$$g(x) \in [-\|C\|, \|C\|], \quad h^*(x) \in (-\infty, 2\|C\|M_w]; \tag{4.7}$$

see (3.11) and Lemma 4.1.

The next lemma establishes fundamental properties of the mappings $g(\cdot)$ and $h^*(\cdot)$.

Lemma 4.2 *With the notation in (4.5)–(4.7) assertions (i)–(iv) below hold.*

- (i) *The mapping $g(\cdot)$ in (4.6) is constant, say $g(x) = g^* \in \mathbb{R}$ for each $x \in S$.*
- (ii) *For each $x \in S$, $e^{\lambda g^* + \lambda h^*(x)} \geq \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h^*(y)} \right]$.*

(iii) For each positive integer n ,

$$ng^* + h^*(x) - 2\|C\|M_w \leq J_n(\pi, x), \quad x \in S, \quad \pi \in \mathcal{P}.$$

(iv) $g^* \leq J_*(\cdot)$.

Proof (i) Notice that (3.8) yields that $g_{\alpha_n}(x) - g_{\alpha_n}(w) = \frac{1 - \alpha_n}{\alpha_n} h_{\alpha_n}(x)$ for every $x \in S$. Taking the limit as n goes to ∞ , (4.6) and (4.7) together yield that $g(x) = g(w)$ for every $x \in S$.

(ii) Let $(x, a) \in \mathbb{K}$ be arbitrary and notice that (3.9) implies that, for each $n \in \mathbb{N}$,

$$e^{\lambda g_{\alpha_n}(x) + \lambda h_{\alpha_n}(x)} \geq e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{\alpha_n}(y)}.$$

Taking the inferior limit as n goes to ∞ in both sides of this inequality, (4.6) and part (i) together imply that

$$\begin{aligned} e^{\lambda g^* + \lambda h^*(x)} &\geq \liminf_{n \rightarrow \infty} e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{\alpha_n}(y)} \\ &\geq e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) \liminf_{n \rightarrow \infty} e^{\lambda h_{\alpha_n}(y)} \end{aligned}$$

where Fatou’s lemma was used to set the second inequality. Thus, (4.6) and the above display lead to

$$e^{\lambda g^* + \lambda h^*(x)} \geq e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h^*(y)}, \quad (x, a) \in \mathbb{K}, \tag{4.8}$$

establishing part (ii).

(iii) An induction argument starting at (4.8) and using the Markov property yields that for every $x \in S, \pi \in \mathcal{P}$ and $n \in \mathbb{N} \setminus \{0\}$,

$$e^{\lambda n g^* + \lambda h^*(x)} \geq E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda h^*(X_{n+1})} \right].$$

From this relation, recalling that $\lambda < 0$ and using (4.7) it follows that

$$e^{\lambda n g^* + \lambda h^*(x)} \geq E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + 2\lambda \|C\| M_w} \right] = e^{\lambda J_n(\pi, x) + 2\lambda \|C\| M_w},$$

where (2.8) was used to set the equality. Therefore, $\lambda n g^* + \lambda h^*(x) \geq \lambda J_n(\pi, x) + 2\lambda \|C\| M_w$, and the conclusion follows, since λ is negative.

(iv) Dividing by n both sides of 4.2 and taking the inferior limit as $n \nearrow \infty$ in the resulting inequality, (2.9) yields that $g^* \leq J(\pi, x)$ for each $x \in S$ and $\pi \in \mathcal{P}$. From this point, (2.10) leads to $g^* \leq J_*(\cdot)$. □

The following result is the final step before proceeding to the proof of the main theorem.

Lemma 4.3 *Given $\alpha \in (0, 1)$, let the policy $f_\alpha \in \mathbb{F}$ be such that (3.7) holds.*

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

$$g_\alpha(x) \geq (1 - \alpha)^2 \sum_{k=1}^{\infty} \alpha^{k-1} J_k(f_\alpha, x).$$

(ii) *Given $\varepsilon > 0$ and $x \in S$, there exists $\tilde{\alpha}_{x,\varepsilon} \in (0, 1)$ such that*

$$g_\alpha + \varepsilon/2 \geq J(f_\alpha, x), \quad \alpha \in (\tilde{\alpha}_{x,\varepsilon}, 1).$$

(iii) $g^* \geq J_*(\cdot)$.

Proof (i) Let $x \in S$ be arbitrary but fixed. Following ideas in Cavazos-Cadena and Salem-Silva (2010), it will be proved by induction that for every positive integer n

$$e^{\lambda V_\alpha(x)} \leq E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right]^{\alpha^n} \prod_{k=1}^n e^{\lambda(1-\alpha)\alpha^{k-1} J_k(f_\alpha, x)}. \quad (4.9)$$

To begin with, recall that the equality $P_x^{f_\alpha}[A_t = f_\alpha(X_t)] = 1$ is always valid, so that the Markov property and (3.7) together yield that, for every $x \in S$ and $n \in \mathbb{N}$,

$$e^{\lambda V_\alpha(X_n)} = E_x^{f_\alpha} \left[e^{\lambda C(X_n, A_n) + \lambda \alpha V_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right], \quad P_x^{f_\alpha}\text{-a. s.}$$

Setting $n = 0$ in this relation and using that $P_x^{f_\alpha}[X_0 = x]$, it follows that

$$\begin{aligned} e^{\lambda V_\alpha(x)} &= E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0) + \lambda \alpha V_\alpha(X_1)} \right] \\ &= E_x^{f_\alpha} \left[\left(e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right)^\alpha \left(e^{\lambda C(X_0, A_0)} \right)^{1-\alpha} \right] \\ &\leq E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right]^\alpha E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0)} \right]^{(1-\alpha)} \\ &= E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right]^\alpha e^{\lambda J_1(f_\alpha, x)(1-\alpha)} \end{aligned}$$

where Hölder's inequality was used in the third step, and the last equality is due to (2.8). This shows that (4.9) holds for $n = 1$. Next, assume that (4.9) is valid for certain positive integer n . Observing that the equality $A_t = f_\alpha(X_t)$ is always valid with probability one under f_α and using that $\sum_{t=0}^{n-1} C(X_t, A_t)$ is \mathcal{F}_n -measurable,

by (2.1), via the Markov property it follows that

$$\begin{aligned} & E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \middle| \mathcal{F}_n \right] \\ &= e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} e^{\lambda V_\alpha(X_n)} \\ &= e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} E_x^{f\alpha} \left[e^{\lambda C(X_n, A_n) + \lambda \alpha V_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \\ &= E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda \alpha V_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right]. \end{aligned}$$

Therefore, via Hölder’s inequality and (2.8) it follows that

$$\begin{aligned} & E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right] \\ &= E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda \alpha V_\alpha(X_{n+1})} \right] \\ &= E_x^{f\alpha} \left[\left(e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right)^\alpha \left(e^{\lambda \sum_{t=0}^n C(X_t, A_t)} \right)^{(1-\alpha)} \right] \\ &\leq E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^\alpha E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t)} \right]^{(1-\alpha)} \\ &= E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^\alpha \left(e^{\lambda J_{n+1}(f_\alpha, x)} \right)^{(1-\alpha)}, \end{aligned}$$

and then

$$\begin{aligned} & E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right]^{\alpha^n} \\ &\leq E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^{\alpha^{n+1}} \left(e^{\lambda J_{n+1}(f_\alpha, x)} \right)^{(1-\alpha)\alpha^n} \\ &= E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^{\alpha^{n+1}} e^{\lambda(1-\alpha)\alpha^n J_{n+1}(f_\alpha, x)}. \end{aligned}$$

Combining this relation with the induction hypothesis, it follows that (4.9) holds with $n + 1$ instead of n . Now, to establish part (i) notice that for $n = 1, 2, 3, \dots$

$$\left| \sum_{t=0}^{n-1} C(X_t, A_t) + V_\alpha(X_n) \right| \leq n \|C\| + \|V_\alpha(\cdot)\| \leq \|C\|(n + (1 - \alpha)^{-1}),$$

so that $E_x^{f\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right] \leq e^{|\lambda| \|C\|(n+(1-\alpha)^{-1})}$, and via (4.9) it follows that

$$e^{\lambda V_\alpha(x)} \leq e^{\alpha^n |\lambda| \|C\|(n+(1-\alpha)^{-1})} \prod_{k=1}^n e^{\lambda(1-\alpha)\alpha^{k-1} J_k(f_\alpha, x)},$$

an inequality that, recalling that $\lambda < 0$, is equivalent to

$$V_\alpha(x) \geq -\alpha^n \|C\| (n + (1 - \alpha)^{-1}) + \sum_{k=1}^n (1 - \alpha) \alpha^{k-1} J_k(f_\alpha, x).$$

Multiplying by $(1 - \alpha)$ both sides of this relation, using (3.8) it follows that

$$g_\alpha(x) \geq -\alpha^n (1 - \alpha) \|C\| (n + (1 - \alpha)^{-1}) + \sum_{k=1}^n (1 - \alpha)^2 \alpha^{k-1} J_k(f_\alpha, x)$$

and the desired conclusion follows taking the limit as n goes to ∞ .

(ii) Let $x \in S$ and $\varepsilon > 0$ be arbitrary and, using (2.9), pick $N_0(x, \varepsilon) \in \mathbb{N}$ such that

$$\frac{1}{k} J_k(f_\alpha, x) \geq J(f_\alpha, x) - \varepsilon/4, \quad k \geq N_0(x, \varepsilon).$$

Thus, observing that $|J(f_\alpha, x)|, k^{-1} |J_k(f_\alpha, x)| \leq \|C\|$, via part (i) it follows that

$$\begin{aligned} g_\alpha(x) &\geq (1 - \alpha)^2 \sum_{k=1}^{\infty} k \alpha^{k-1} \frac{J_k(f_\alpha, x)}{k} \\ &= J(f_\alpha, x) + (1 - \alpha)^2 \sum_{k=1}^{\infty} k \alpha^{k-1} \left(\frac{1}{k} J_k(f_\alpha, x) - J(f_\alpha, x) \right) \\ &\geq J(f_\alpha, x) + (1 - \alpha)^2 \sum_{k=1}^{N_0(x, \varepsilon)-1} k \alpha^{k-1} \left(\frac{1}{k} J_k(f_\alpha, x) - J(f_\alpha, x) \right) - \varepsilon/4 \\ &\geq J(f_\alpha, x) - 2(1 - \alpha)^2 \|C\| \sum_{k=1}^{N_0(x, \varepsilon)-1} k \alpha^{k-1} - \varepsilon/4. \end{aligned}$$

where the previous display was used to set the first inequality. Finally, select $\tilde{\alpha}_{x, \varepsilon}$ such that $(1 - \alpha)^2 \sum_{k=1}^{N_0(x, \varepsilon)-1} k \alpha^{k-1} \leq \varepsilon(8\|C\| + 1)^{-1}$ when $\alpha \in (\tilde{\alpha}_{x, \varepsilon}, 1)$ to conclude that

$$g_\alpha(x) \geq J(f_\alpha, x) - \varepsilon/2, \quad \alpha \in (\tilde{\alpha}_{x, \varepsilon}, 1),$$

completing the proof of part (ii).

(iii) Let $x \in S$ be arbitrary. Given $\varepsilon > 0$, let $\tilde{\alpha}_{x, \varepsilon} \in (0, 1)$ be as in part (ii) and observe that (4.5) yields that there exists $\tilde{N}(x, \varepsilon) \in \mathbb{N}$ such that $\alpha_n > \tilde{\alpha}_{x, \varepsilon}$ if $n > \tilde{N}(x, \varepsilon)$, and in this case (4.3) implies that $g_{\alpha_n}(x) \geq J(f_{\alpha_n}, x) - \varepsilon/2$, so that

$$g_{\alpha_n}(x) \geq J_*(x) - \varepsilon/2, \quad n > \tilde{N}(x, \varepsilon).$$

Taking the limit as n goes to ∞ , this relation leads to $g^* \geq J_*(x) - \varepsilon/2$, and the conclusion follows, since $\varepsilon > 0$ is arbitrary. \square

5 Proof of the main result

After the preliminaries in the previous section, the main conclusions of the paper can be established as follows.

Proof of Theorem 3.1 Let $\{\alpha_n\}_{n \in \mathbb{N}}$ be an arbitrary sequence satisfying (4.5) and, as before, taking a subsequence, if necessary, without loss of generality assume that (4.6) holds, so that $\lim_{k \rightarrow \infty} g_{\alpha_k}(\cdot) = g^* \in \mathbb{R}$, by Lemma 4.2(i).

- (i) Combining Lemma 4.2(iv) and Lemma 4.3(iii) it follows that $J_*(\cdot) = g^* = \lim_{n \rightarrow \infty} g_{\alpha_n}(x)$ for every $x \in S$. Thus, since the sequence $\{\alpha_n\}$ satisfying (4.5) is arbitrary, it follows that $\lim_{\alpha \nearrow 1} g_\alpha(\cdot) = J_*(\cdot) = g^*$.
- (ii) Let $x \in S$ be arbitrary but fixed. Given $\varepsilon > 0$, using part (i) select $\hat{\alpha}_{x,\varepsilon} \in (0, 1)$ such that

$$g_\alpha(x) < g^* + \varepsilon/2, \quad \alpha \in (\hat{\alpha}_{x,\varepsilon}, 1).$$

Setting $\alpha_{x,\varepsilon} = \max\{\hat{\alpha}_{x,\varepsilon}, \tilde{\alpha}_{x,\varepsilon}\}$, this last display and Lemma 4.3 (ii) together yield that (3.12) holds.

6 Conclusion

In this work, Markov decision chains on a denumerable state space were studied. It was assumed that the performance of a decision policy is measured by the average criterion as perceived by a risk-seeking controller with constant risk-sensitivity. Under conditions ensuring that the optimal average cost is constant, but *not* that the optimality equation admits a solution, the problems of approximating the optimal average cost, and determining a nearly optimal policy via the family of fixed points of contractive operators were studied. The results in this direction, which are stated in Theorem 3.1, provide an extension to the present framework of the classical *discounted approach* in the theory of Markov decision chains endowed with the risk-neutral average index. On the other hand, extending the conclusions in Theorem 3.1 to more general contexts, including unbounded costs or more general state space, seems to be an interesting problem.

Acknowledgements 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. H. Cruz-Suárez and G. Portillo-Ramírez 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.

Funding This research is part of the Ph D research work of the first author, and no special resources were received to perform this work.

Data availability This work does not require any special data.

Declarations

Conflict of interest Hereby the authors declare that every result already published in the literature that is used in this work has been properly recognized. The main research topic was suggested to the first author by the second and third authors, and the first author performed his research work under the guidance of the second and third author.

References

- Arapostathis A, Borkar VS, Fernández-Gaucherand E, Ghosh MK, Marcus SI (1993) Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J Control Optim* 31(2):282–334. <https://doi.org/10.1137/0331018>
- Barz C, Waldmann KH (2007) Risk-sensitive capacity control in revenue management. *Math Methods Oper Res* 65(3):565–579. [https://doi.org/10.1016/S0304-4149\(00\)00032-6](https://doi.org/10.1016/S0304-4149(00)00032-6)
- Basu A, Ghosh MK (2014) Zero-sum risk-sensitive stochastic games on a countable state space. *Stoch Process Appl* 124(1):961–983. <https://doi.org/10.1016/j.spa.2013.09.009>
- Bäuerle N, Rieder U (2011) Markov decision processes with applications to finance. Springer, New York
- Bäuerle N, Rieder U (2014) More risk-sensitive Markov decision processes. *Math Oper Res* 39(1):105–120. <https://doi.org/10.1287/moor.2013.0601>
- Borkar VS, Meyn SP (2002) Risk-sensitive optimal control for Markov decision process with monotone cost. *Math Oper Res* 27(1):192–209. <https://doi.org/10.1287/moor.27.1.192.334>
- Cavazos-Cadena R (2009) Solutions of the average cost optimality equation for finite Markov decision chains: risk-sensitive and risk-neutral criteria. *Math Method Oper Res* 70:541–566
- Cavazos-Cadena R (2018) Characterization of the optimal risk-sensitive average cost in denumerable Markov decision chains. *Math Oper Res* 43(3):1025–1050. <https://doi.org/10.1287/moor.2017.0893>
- Cavazos-Cadena R, Fernández-Gaucherand E (1999) Controlled Markov chains with risk-sensitive criteria: average cost, optimality equations, and optimal solutions. *Math Method Oper Res* 49:299–324
- Cavazos-Cadena R, Salem-Silva F (2010) The discounted method and equivalence of average criteria for risk-sensitive Markov decision processes on Borel spaces. *Appl Math Optim* 61(2):167–190
- Di Masi GB, Stettner L (1999) Risk-sensitive control of discrete time Markov processes with infinite horizon. *SIAM J Control Optim* 38(1):61–78. <https://doi.org/10.1137/S0363012997320614>
- Di Masi GB, Stettner L (2000) Infinite horizon risk sensitive control of discrete time Markov processes with small risk. *Syst Control Lett* 40:15–20. [https://doi.org/10.1016/S0167-6911\(99\)00118-8](https://doi.org/10.1016/S0167-6911(99)00118-8)
- Di Masi GB, Stettner L (2007) Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM J Control Optim* 46(1):231–252. <https://doi.org/10.1137/040618631>
- Hernández-Hernández D, Marcus SI (1996) Risk-sensitive control of Markov processes in countable state space. *Syst Control Lett* 29:147–155
- Hernández-Lerma O (1989) Adaptive Markov control processes. Springer, New York
- Howard RA, Matheson JE (1972) Risk-sensitive Markov decision processes. *Manag Sci* 18(7):356–369
- Jaśkiewicz A (2007) Average optimality for risk sensitive control with general state space. *Ann Appl Probab* 17(2):654–675. <https://doi.org/10.1214/105051606000000790>
- Jaśkiewicz A, Nowak AS (2014) Stationary Markov perfect equilibria in risk sensitive stochastic overlapping generations models. *J Econ Theory* 151:411–447. <https://doi.org/10.1016/j.jet.2014.01.005>
- Pitera M, Stettner L (2016) Long run risk sensitive portfolio with general factors. *Math Methods Oper Res* 82(2):265–293. <https://doi.org/10.1007/s00186-015-0528-7>
- Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, New York
- Saucedo-Zul J, Cavazos-Cadena R, Cruz-Suárez H (2020) A discounted approach in communicating average Markov decision chains under risk-aversion. *J Optim Theory Appl* 187:585–606
- Shen Y, Stannat W, Obermayer K (2013) Risk-sensitive Markov control processes. *SIAM J Control Optim* 51(5):3652–3672. <https://doi.org/10.1137/120899005>
- Sladký K (2008) Growth rates and average optimality in risk-sensitive Markov decision chains. *Kybernetika* 44(2):205–226
- Sladký K (2018) Risk-sensitive average optimality in Markov decision processes. *Kybernetika* 54(6):1218–1230. <https://doi.org/10.14736/kyb-2018-6-1218>

Stettner L (1999) Risk sensitive portfolio optimization. *Math Meth Oper Res* 50(3):463–474. <https://doi.org/10.1007/s001860050081>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.