

Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions*

Rolando Cavazos-Cadena^{1,**}, Emmanuel Fernández-Gaucherand²

¹Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo COAH 25315, México

²Systems and Industrial Engineering Department, The University of Arizona, Tuscon, AZ, 85721-0020, USA

Abstract. We study controlled Markov chains with denumerable state space and bounded costs per stage. A (long-run) risk-sensitive average cost criterion, associated to an exponential utility function with a constant risk sensitivity coefficient, is used as a performance measure. The main assumption on the probabilistic structure of the model is that the transition law satisfies a simultaneous *Doebelin condition*. Working within this framework, the main results obtained can be summarized as follows: If the constant risk-sensitivity coefficient is *small enough*, then an associated optimality equation has a bounded solution with a *constant value* for the optimal risk-sensitive average cost; in addition, under further standard continuity-compactness assumptions, optimal stationary policies are obtained. However, it is also shown that the above conclusions *fail* to hold, in general, for *large enough* values of the risk-sensitivity coefficient. Our results therefore disprove previous claims on this topic. Also of importance is the fact that our developments are very much self-contained and employ only basic probabilistic and analysis principles.

Key words: Controlled Markov chains, exponential utility function, constant risk sensitivity, simultaneous Doebelin condition, bounded solutions to the risk-sensitive optimality equation, constant average cost.

* This work was partially supported by a U.S.-México Collaborative Program, under grants from the National Science Foundation (NSF-INT 9602939), and the Consejo Nacional de Ciencia y Tecnología (CONACyT) (No. E 120.3336).

** The support of the PSF Organization under Grant No. 200-350-97-04 is deeply acknowledged by the first author.

Manuscript received: March 1998/final version received: July 1998

1 Introduction

In this paper, discrete-time controlled Markov chains (CMC's) are studied within the following modeling framework: (i) the state space is denumerable, (ii) the transition law satisfies the strong form of the simultaneous Doeblin condition [1, 2, 3, 11, 17, 18, 20] (see Assumption 2.2 below), and (iii) the one-stage cost function is bounded. An exponential utility function U with (non-null) constant risk sensitivity [12, 14, 21–23] is used to assess the value of costs incurred, so that when facing a random cost \mathcal{C} , the controller can interchange the possibility of incurring \mathcal{C} with the opportunity of paying the constant cost c that satisfies $U(c) = E[U(\mathcal{C})]$; such a value c is called the *certain equivalent* of \mathcal{C} with respect to U . The performance index of a control policy π is the *risk-sensitive (long-run) average cost criterion*, which is constructed by the following two-step procedure: first, the certain equivalent of the total cost incurred by π up to a positive time n is computed, and second, the limit (superior) of the certain equivalent per unit time is calculated.

The study of controlled stochastic dynamical systems with risk-sensitive criteria can be traced back, at least, to [12] and [13]. Particularly, in [12] the case of CMC's with finite state and action spaces was considered. Under a “primitiveness” assumption on the transition law, an equation for the value attained by stationary policies was obtained via matrix analysis, and the convergence of a policy improvement algorithm was established. Recently, there has been a rekindled interest on controlled stochastic processes endowed with risk-sensitive criteria [5–8, 15, 19, 21–23]. The main purpose of this paper is to study the existence of (bounded) solutions to the risk-sensitive average cost optimality equation, with a constant value of the optimal average cost. Under suitable continuity-compactness conditions, this optimality equation yields an optimal stationary policy. The risk-neutral version of this problem has been widely studied in the literature, and a fairly complete theory for this case is now available; see [1, 3, 4, 11, 20].

The *main* result of the paper, stated below as Theorem 3.1, guarantees the existence of a bounded solution to the optimality equation when the risk sensitivity coefficient λ is *small enough*. On the other hand, a detailed example is given in Section 3 showing that such a conclusion *fails* to hold in general for arbitrary values of λ ; see Proposition 3.1 and Remark 3.1 below. Our results therefore disprove previous claims on this topic, e.g., see [8, 14].

Other recent efforts in this topic have relied on the use of associated stochastic games, e.g., see [8, 14, 15, 19]. In contrast, our results are obtained by fairly self-contained arguments which use only basic probability and analysis principles. At the same time, our analysis differs from the usual CMC methodology in the following aspect: When the risk-neutral average cost is studied, the existence of (bounded) solutions of the optimality equation is usually established either using contractive mappings or via the so-called vanishing discount approach [1, 9, 17, 18], whereas the proof of our main results rely on the study of a parameterized expected-total cost problem with a stopping time, and the result is obtained by an appropriate selection of the parameter; this approach allows to include *both* the risk-averse and risk-seeking cases.

The organization of the paper is as follows: In Section 2 the decision model is introduced and the main result is stated in Section 3 in the form of Theorem 3.1; also, an example is used to show that the conclusions in Theorem 3.1 can not be extended to arbitrary risk sensitivity coefficients. Next, Section 4 and 5

contain the preliminaries that will be used to establish Theorem 3.1 in Section 6. Finally, the paper concludes in Section 7 with some brief comments, followed by a bibliography.

Notation. Throughout the remainder \mathbb{R} and \mathbb{N}_0 stand for the set of real numbers and nonnegative integers, respectively. Given a nonempty set S , $\mathcal{M}_b(S)$ denotes the space of all (measurable) real-valued bounded functions defined on S , i.e., $\mathcal{M}_b(S) := \{C : S \rightarrow \mathbb{R} \mid \|C\| < \infty\}$, where $\|C\| := \sup_w |C(w)|$ is the supremum norm of C . On the other hand, δ_{xy} denotes the Kronecker delta function, that is, $\delta_{xy} = 1$ (resp. 0) if $x = y$ (resp. $x \neq y$), and $a \wedge b := \min\{a, b\}$ for $a, b \in \mathbb{R}$. Finally, for an event W , the corresponding indicator function is denoted by $\mathcal{I}[W]$. As usual, all relations involving conditional expectation are supposed to hold true almost everywhere with respect to the underlying probability measure without explicit reference.

2 The decision model

Following standard notation [1], [9], let $\langle S, A, C, P \rangle$ denote the CMC model, where the state space S is a denumerable set, the (nonempty) separable metric space A is the control (or action) set, $C : S \times A \rightarrow \mathbb{R}$ is the one-stage cost function and $P = [p_{xy}(\cdot)]$ is the controlled transition law. At each time $t \in \mathbb{N}_0$ the state of a dynamical system is observed, say $X_t = x \in S$, and an action $A_t = a \in A$ is chosen. Then a cost $C(x, a)$ is incurred 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 $p_{xy}(a)$; this is the Markov property of the decision model.

Remark 2.1. Notice that it is assumed that every $a \in A$ is an admissible action at each state; however, as noted in [2], this condition does not imply any loss of generality.

The discussion below requires the following basic condition.

Assumption 2.0. $C \in \mathcal{M}_b(S \times A)$, and for each $x, y \in S$, $a \mapsto C(x, a)$ and $a \mapsto p_{xy}(a)$ are measurable mappings on A .

On the other hand, the results on existence of optimal stationary policies will be derived under the following stronger assumption.

Assumption 2.1. (i) The action set A is compact. (ii) For each $x, y \in S$, $a \mapsto C(x, a)$ and $a \mapsto p_{xy}(a)$ are continuous mappings on A .

Policies. For each $t \in \mathbb{N}_0$ the space of histories up to time t is recursively defined by $H_0 := S$ and $H_t := H_{t-1} \times A \times S$; a generic element of H_t is denoted by $h_t = (x_0, a_0, \dots, x_{t-1}, a_{t-1}, x_t)$, where $x_i \in S$ and $A_i \in A$. A control policy is a sequence $\pi = \{\pi_t\}$ where each π_t is a stochastic kernel on A given H_t . That is, for each $h_t \in H_t$, $\pi_t(\cdot | h_t)$ is a probability measure on A ; for each Borel subset $B \subset A$, the number $\pi_t(B | h_t)$ is the probability of choosing an action $A_t \in B$, and it is assumed that $\pi_t(B | \cdot)$ is a measurable mapping on H_t . Throughout the remainder Π denotes the class of all policies. Given the policy

π being used and the initial state $X_0 = x$, the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined [1, 9, 10, 17]; such a distribution is denoted by P_x^π whereas E_x^π stands for the corresponding expectation operator. Let $\mathbb{F} := \prod_{x \in S} A$, so that \mathbb{F} consists of all functions $f : S \rightarrow A$. A policy π is stationary if there exists $f \in \mathbb{F}$ such that, under π , at each time t the action applied is $A_t = f(X_t)$. The class of stationary policies is naturally identified with \mathbb{F} , and with this convention $\mathbb{F} \subset \Pi$. Notice, finally, that under the action of each stationary policy, the state process $\{X_t\}$ is a Markov chain with stationary transition probabilities.

Utility function. For each $\lambda \in \mathbb{R}$ define $U_\lambda : \mathbb{R} \rightarrow \mathbb{R}$, the (exponential) *utility function* with (constant) risk sensitivity λ , as follows: For $x \in \mathbb{R}$,

$$U_\lambda(x) := \begin{cases} \text{sign}(\lambda)e^{\lambda x}, & \lambda \neq 0, \\ x, & \lambda = 0; \end{cases} \quad (2.1)$$

notice that each function $U_\lambda(\cdot)$ is increasing. A decision maker (DM) with utility function given by (2.1) exhibits *constant risk aversion* as manifested by λ , and the DM will be: *risk averse* if $\lambda > 0$, *risk seeking* if $\lambda < 0$, and *neutral to risk* if $\lambda = 0$; see [12, 14, 21–23].

For a (bounded) random variable Y , the corresponding certain equivalent $E(\lambda, Y)$ with respect to U_λ is implicitly defined by

$$U_\lambda(E(\lambda, Y)) = E[U_\lambda(Y)], \quad (2.2)$$

so that a controller with risk sensitivity λ is indifferent between incurring the random cost Y or paying the corresponding certain equivalent for sure. Observe now that (2.1) and (2.2) together yield that

$$E(\lambda, Y) = \begin{cases} \frac{1}{\lambda} \log(E[e^{\lambda Y}]), & \lambda \neq 0 \\ E[Y] & \lambda = 0. \end{cases} \quad (2.3)$$

Using Jensen's inequality, it follows that $E(\lambda, Y) > E[Y]$ when $\lambda > 0$ and Y is non-constant, i.e., the DM is risk averse in that the certain quantity $E(\lambda, Y)$ is preferred over the expected value of the random (uncertain) cost Y .

Remark 2.2. For a random variable Y let

$$\text{ess inf}(Y) := \inf\{m \mid P[Y < m] > 0\} \quad \text{and}$$

$$\text{ess sup}(Y) := \sup\{m \mid P[Y > m] > 0\}$$

be the essential infimum and essential supremum of Y , respectively. With this notation $\text{ess inf}(Y) \leq Y \leq \text{ess sup}(Y)$ (almost surely), and then, since $U_\lambda(\cdot)$ is increasing, $U_\lambda(\text{ess inf}(Y)) \leq E[U_\lambda(Y)] \leq U_\lambda(\text{ess sup}(Y))$, and the definition of the certain equivalent in (2.2) yields that $\text{ess inf}(Y) \leq E(\lambda, Y) \leq \text{ess sup}(Y)$.

Performance Index. Let $n \in \mathbb{N}_0$, then under the action of $\pi \in \Pi$ and given $X_0 = x \in S$, $J_n(\lambda, \pi, x)$ denotes the certain equivalent of the total cost incurred up to time n with respect to U_λ , i.e.,

$$J_n(\lambda, \pi, x) = \begin{cases} \frac{1}{\lambda} \log(E_x^\pi [e^{\lambda \sum_{t=0}^n C(X_t, A_t)}]), & \lambda \neq 0 \\ E_x^\pi \left[\sum_{t=0}^n C(X_t, A_t) \right], & \lambda = 0, \end{cases} \tag{2.4}$$

whereas the long-run λ -sensitive average cost under π starting at x is defined by

$$J(\lambda, \pi, x) = \limsup_{n \rightarrow \infty} \frac{1}{n+1} J_n(\lambda, \pi, x). \tag{2.5}$$

The optimal λ -sensitive average cost at state x is given by

$$J^*(\lambda, x) = \inf_{\pi} J(\lambda, \pi, x), \tag{2.6}$$

and a policy $\pi^* \in \Pi$ is λ -average optimal (λ -AO) if $J^*(\lambda, x) = J(\lambda, \pi^*, x)$, for every $x \in S$.

Remark 2.3. Note that if $a \leq C(\cdot, \cdot) \leq b$, for $a, b \in \mathbb{R}$, then $(n+1)a \leq J_n(\lambda, \pi, x) \leq (n+1)b$ (see Remark 2.2), and thus $a \leq J(\lambda, \pi, x) \leq b$. This point will be used repeatedly in the sequel.

From the literature on the risk-neutral average cost criterion (i.e. $\lambda = 0$), it is well-known that a ‘‘communicating’’ condition is necessary in order to have that the optimal average cost is independent of the initial state, and that a strong recurrence condition is required for the existence of a bounded solution to the average cost optimality equation; see [1, 3, 4, 9, 17, 20]. To study the risk-sensitive average criterion, the following (Doebelin) condition will be used throughout the remainder of the paper; see also [1, 9, 11, 17, 18, 20].

Assumption 2.2. (Simultaneous Doebelin Condition). There exist a state $z \in S$ and a positive integer K such that

$$E_x^f [T] \leq K, \quad \text{for all } x \in S \quad \text{and} \quad f \in \mathbb{F}, \tag{2.7}$$

where T is the first passage time to state z , i.e.,

$$T := \min\{n > 0 \mid X_n = z\}. \tag{2.8}$$

Note that, with no loss in generality, K is restricted to be an integer in Assumption 2.2, which will ease the subsequent presentation. The following lemma summarizes some well-known consequences of Assumptions 2.1 and 2.2 for the risk-neutral average cost criterion.

Lemma 2.1. *Suppose that Assumptions 2.1 and 2.2 hold true. In this case there exists a constant $g \in \mathbb{R}$ and $h : S \rightarrow \mathbb{R}$ such that*

- (i) $\|h\| < \infty$;
- (ii) *The risk-neutral optimal average cost is constant and equal to g , that is $J^*(0, x) = g$ for all $x \in S$ (see (2.4)–(2-6)).*
- (iii) *The pair $(g, h(\cdot))$ satisfies the (risk-neutral) average cost optimality equation:*

$$g + h(x) = \inf_{a \in A} \left[C(x, a) + \sum_y p_{xy}(a)h(y) \right], \quad x \in S. \quad (2.9)$$

Moreover,

- (iv) *For each $x \in S$ the term within brackets in (2.9) is a continuous function of $a \in A$. Consequently, (2.9) has a minimizer $f(x) \in A$, and the corresponding policy $f \in \mathbb{F}$ is (risk-neutral) average optimal.*

A proof of the result above can be found in, e.g., [1, 9, 17, 18]. The remainder of this paper presents extensions of Lemmas 2.1 and 2.2 to the risk-sensitive context. In contrast to claims in [8] (see also [14]), it will be shown that the risk-sensitive counterpart to Lemma 2.1 *does not hold*, in general, for arbitrary $\lambda \in \mathbb{R}$.

3 Solutions to the optimality equation

As already mentioned, the main purpose of this paper is to study the existence of bounded solutions to the λ -average cost optimality equation (λ -ACOE) associated to a risk-sensitive average cost criterion. We show via a detailed example that, in general, the λ -ACOE *does not* admit a bounded solution with constant optimal average cost for arbitrary values of λ . Furthermore, subsequently we present a result, Theorem 3.1 below, showing that the λ -ACOE does admit such solutions *whenever the risk sensitivity coefficient λ is sufficiently small*. Hence, even under the strong recurrence assumptions being employed, it is shown that the risk-sensitive average cost problem is “well behaved” only for situations that are close enough to the risk-neutral, or standard, situation.

Example 3.1. Let the state space be $S = \{0, 1\}$ and suppose that the control space is a singleton: $A = \{a\}$, so that the space of policies also contains just one element. Define the transition law by

$$p_{10}(a) = 1 - p_{11}(a) = q, \quad p_{00}(a) = 1$$

where $q \in (0, 1)$ is a given number. Next set $z = 0 \in S$ and observe that $P_z[T = 1] = 1$, whereas $P_1[T = k] = q(1 - q)^{k-1}$, for every positive integer k . Therefore,

$$E_z[T] = 1 \quad \text{and} \quad E_1[T] = \frac{1}{q},$$

and it is clear that Assumption 2.2 is satisfied in this example. □

Proposition 3.1. *Let λ with $|\lambda| > -\log(1 - q)$ be fixed, and select the one-stage cost function as $C(1, a) = \text{sign}(\lambda)$ and $C(0, a) = 0$; denote $C_\lambda(x) := C(x, a)$. Then, (i)–(ii) below hold true:*

- (i) *The λ -average cost function – i.e., $J^*(\lambda, \cdot)$ – corresponding to $C_\lambda(\cdot)$ is **not** constant. Consequently,*
- (ii) *The λ -ACOE associated to the the risk sensitivity coefficient λ **does not** have a solution leading to a constant average cost, i.e., there are **no** real numbers $g_\lambda, h_\lambda(1)$, and $h_\lambda(0)$ for which*

$$e^{\lambda g_\lambda + h_\lambda(x)} = E_x[e^{\lambda C_\lambda(X_0) + h_\lambda(X_1)}], \quad x \in S.$$

Remark 3.1. (i) Assumptions 2.0–2.2 are satisfied in Example 3.1 but, as given by Proposition 3.1(i), the average cost calculated by a DM with sufficiently large risk sensitivity is *not necessarily* constant; this feature must be contrasted with the risk-neutral case, in which the (optimal) average cost is independent of the initial state and the (risk-neutral) average cost (Poisson) equation does have a solution.

(ii) Proposition 3.1 explicitly disproves claims in, e.g., [8] (see also [14]), where it was stated that under Assumptions 2.1 and 2.2 the λ -ACOE has a bounded solution for every $\lambda > 0$, with an optimal average cost independent of the initial state.

Proof of Proposition 3.1. Using that a null cost is incurred at $z = 0$, which is an absorbing state, it is clear that $J^*(\lambda, 0) = 0$, whereas starting at $X_0 = 1$,

$$\sum_{t=0}^n C_\lambda(X_t) = \text{sign}(\lambda)(T \wedge (n + 1)),$$

since $C_\lambda(1) = \text{sign}(\lambda)$ and $X_t = 1$ for $t < T$; see (2.8). Thus,

$$e^{\lambda J_n(\lambda, 1)} = E_1[e^{\lambda \sum_{t=0}^n C_\lambda(X_t)}] = E_1[e^{\lambda \text{sign}(\lambda)(T \wedge (n+1))}] = E_1[e^{|\lambda|(T \wedge (n+1))}],$$

and thus

$$\begin{aligned} e^{\lambda J_n(\lambda, 1)} &= \sum_{k=1}^n e^{|\lambda|k} P_1[T = k] + e^{|\lambda|(n+1)} P_1[T > n] \\ &= \sum_{k=1}^n e^{|\lambda|k} q(1 - q)^{k-1} + e^{|\lambda|(n+1)}(1 - q)^n \\ &= e^{|\lambda|} q \sum_{k=0}^{n-1} e^{|\lambda|k} (1 - q)^k + e^{|\lambda|(n+1)}(1 - q)^n \\ &= e^{|\lambda|} q \frac{((1 - q)e^{|\lambda|})^n - 1}{(1 - q)e^{|\lambda|} - 1} + e^{|\lambda|(n+1)}(1 - q)^n \\ &= a(1 - q)e^{|\lambda|} - b \end{aligned}$$

for appropriate positive constants a, b depending on λ , but not on n . Observe that the condition $|\lambda| > -\log(1 - q)$ is equivalent to $(1 - q)e^{|\lambda|} > 1$, and hence it follows that

$$\lim_{n \rightarrow \infty} e^{\lambda J_n(\lambda, 1)/(n+1)} = \lim_{n \rightarrow \infty} (a((1 - q)e^{|\lambda|})^n - b)^{1/(n+1)} = (1 - q)e^{|\lambda|},$$

which is equivalent to

$$\text{sign}(\lambda) \lim_{n \rightarrow \infty} \frac{1}{n + 1} J_n(\lambda, 1) = 1 + \frac{\log(1 - q)}{|\lambda|} > 0$$

so that $J^*(\lambda, 1) \neq 0 = J^*(\lambda, 0)$, and thus the long-run average cost function corresponding to the risk sensitivity coefficient λ with $|\lambda| > -\log(1 - q)$ is not constant, establishing part (i). On the other hand, the existence of a solution of the stated (Poisson) optimality equation implies that the average cost function is constant; this can be obtained from the verification theorem in [8], or from the proof of Theorem 3.1(v) to be presented in Section 6. Therefore, part (ii) follows from part (i). \square

On the positive side, Theorem 3.1 below establishes precise bounds on λ such that the λ -ACOE admits a bounded solution with constant optimal average cost. First, useful notation employed in the sequel is defined.

Definition 3.1. *Suppose that Assumptions 2.0 and 2.2 hold true, and let the positive integer K be as in Assumption 2.2. Define*

$$\beta = \left(\frac{K}{K + 1} \right)^{1/(K+1)},$$

and note that $\beta < 1$.

Theorem 3.1. *Suppose that Assumptions 2.0 and 2.2 hold true and set*

$$\mu := -\log(\beta) = \frac{\log(K + 1) - \log(K)}{K + 1}, \tag{3.1}$$

where the integer K is as in (2.7). If the risk sensitivity coefficient λ is such that

$$\lambda \in \left(-\frac{\mu}{2\|C\|}, \frac{\mu}{2\|C\|} \right) \setminus \{0\},$$

then there exist a constant $g_\lambda \in \mathbb{R}$ and a function $h_\lambda : S \rightarrow \mathbb{R}$ satisfying the following:

- (i) h_λ is bounded.
- (ii) The pair $(g_\lambda, h_\lambda(\cdot))$ satisfies the λ -ACOE:

$$\text{sign}(\lambda)e^{\lambda g_\lambda + h_\lambda(x)} = \min_{a \in A} \left[\text{sign}(\lambda)e^{\lambda C(x, a)} \sum_y p_{xy}(a)e^{h_\lambda(y)} \right], \quad x \in S. \tag{3.2}$$

- (iii) $J^*(\lambda, x) = g_\lambda$ for each $x \in S$. Moreover,
- (iv) The pair $(g_\lambda, h_\lambda(\cdot)) \in \mathbb{R} \times \mathcal{M}_b(S)$ in (3.2) is unique whenever h_λ satisfies $h_\lambda(z) = 0$.
- (v) If, additionally, Assumption 2.1 is valid, then for every $x \in S$, the term in brackets in the right-hand side of (3.2) is a continuous function on A ; thus, it has a minimizer $f^*(x) \in A$ and the corresponding policy $f^* \in \mathbb{F}$ is λ -AO.

The proof of this result will be presented in Section 6. In Section 4 some necessary technical preliminaries are presented, and in Section 5 an auxiliary stopping problem with total cost criterion is introduced. The key idea behind the proof of Theorem 3.1 is to relate the average cost due to a policy π to the expected total cost incurred between visits to the recurrent state z , and the corresponding relative costs $h(x)$ associated with the initial value x of the state are constructed as the total cost up to the first visit to z ; see also [1, p. 315] for the risk-neutral case.

4 Technical preliminaries

In this section several technical preliminaries are introduced. These results are needed in order to introduce, in the following section, the auxiliary expected total-cost problems used to construct the solutions to the λ -ACOE in Theorem 3.1. The first lemma below extends the inequality in Assumption 2.2 to the class of *all* policies. Versions of this result can be found in the literature of CMC with risk-neutral average cost criterion, e.g., [11], but they are derived under the continuity-compactness condition in Assumption 2.1, instead of the weaker Assumption 2.0. For this reason, as well as for the sake of completeness, a detailed proof is provided in the Appendix. Throughout the remainder of this section, Assumptions 2.0 and 2.2 will be supposed to hold true even without explicit reference.

Lemma 4.1. *Let the integer K be as in Assumption 2.2. In this case, for every policy $\pi \in \Pi$ and $x \in S$,*

$$E_x^\pi[T] \leq K.$$

Lemma 4.1 allows to ensure a geometric decay on the tails of the distribution of the first passage time T in (2.3), under every policy $\pi \in \Pi$, as given by the following result.

Lemma 4.2. *Let β be as in Definition 3.1. For every $\pi \in \Pi$ and $n \in \mathbb{N}$,*

$$P_x^\pi[T \geq n] \leq \beta^{n-K}, \quad n \in \mathbb{N}. \tag{4.1}$$

Proof. Notice that (4.1) is valid for $n = 0, 1, \dots, K$, since $\beta < 1$, and that Markov’s inequality yield $P_x^\pi[T \geq K + 1] \leq E_x^\pi[T]/(K + 1) \leq K/(K + 1) < \beta$, so that (4.1) is also valid for $n = K + 1$. The proof is completed by induction as follows. Suppose that, for every policy $\pi \in \Pi$ and $x \in S$, (4.1) occurs for all positive integers $n < m$, where $m > K + 1$. In this case, the Markov

property and the definition of T in (2.8) together imply that

$$\begin{aligned} P_x^\pi[T \geq m | X_i = x_i, i = 1, 2, \dots, K+1] &= P_x^\pi[X_i \neq z, i = 1, 2, \dots, m-1 | X_i = x_i, i = 1, \dots, K+1] \\ &= \begin{cases} P_{x_{K+1}}^{\pi'}[X_i \neq z, i = 1, \dots, m - (K+1) - 1] \\ = P_{x_{K+1}}^{\pi'}[T \geq m - (K+1)], & \text{if } x_i \neq z \forall i \leq K+1, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

where π' is the shifted policy defined by $\pi'_i(\cdot | h_t) = \pi_{t+K+1}(\cdot | x_0, a_0, x_1, \dots, x_K, a_K, h_t)$. Therefore, the induction hypothesis and Markov's inequality yield

$$P_x^\pi[T \geq m | X_1, \dots, X_{K+1}] \leq \mathcal{J}[T > K+1] \beta^{m-(K+1)-K},$$

and then

$$\begin{aligned} P_x^\pi[T \geq m] &\leq P_x^\pi[T > K+1] \beta^{m-(K+1)-K} \\ &\leq \frac{E_x^\pi[T]}{K+1} \beta^{m-(K+1)-K} \\ &\leq \frac{K}{K+1} \beta^{m-(K+1)-K} \\ &= \beta^{K+1} \beta^{m-(K+1)-K} \\ &= \beta^{m-K} \end{aligned}$$

so that (4.1) is also valid for $n = m$ and the induction proof is complete. \square

Theorem 4.1. *Suppose that Assumptions 2.0 and 2.2 hold true, let the positive integer K be as in Assumption 2.2 and let β be as in Definition 3.1. Then the following inequality holds true for every $\pi \in \Pi$ and $x \in S$:*

$$E_x^\pi[e^{\lambda T}] \leq \frac{1}{\beta^K(1 - \beta e^\lambda)} \quad \text{if } \lambda < -\log(\beta).$$

Proof. Notice that $E_x^\pi[e^{\lambda T}] = \sum_{t=0}^{\infty} e^{\lambda t} P_x^\pi[T = t] \leq \sum_{t=0}^{\infty} e^{\lambda t} P_x^\pi[T \geq t]$, so that Lemma 4.2 implies that

$$E_x^\pi[e^{\lambda T}] \leq \beta^{-K} \sum_{t=0}^{\infty} e^{\lambda t} \beta^t = \beta^{-K} \sum_{t=0}^{\infty} (\beta e^\lambda)^t$$

and then $E_x^\pi[e^{\lambda T}] \leq \beta^{-K}(1 - \beta e^\lambda)^{-1} < \infty$, if $\beta e^\lambda < 1$, condition that is clearly equivalent to $\lambda < -\log(\beta)$. \square

5 Auxiliary stopping expected total-cost problems

This section considers an auxiliary model and two related stopping problems, endowed with risk-sensitive expected total-cost criteria. Solutions to the λ -ACOE associated with our original model, see Theorem 3.1, are constructed based on solutions to these auxiliary stopping problems. Essentially, the system is allowed to run until it reaches state z in a positive time, and at that moment it is stopped without incurring any cost. The auxiliary model employed differs from the original model only in the selection of the one-stage cost function, taken as the *deviation cost* $\tilde{C}(\cdot, \cdot) := C(\cdot, \cdot) - g$, where “the parameter” g belongs to the interval $[-\|C\|, \|C\|]$. The performance index of a control policy is the expectation of $\text{sign}(\lambda)U_\lambda(\sum_{t=0}^{T-1}(C(X_t, A_t) - g))$; both maximization and minimization of this criterion will be considered. Formally, for each $\lambda \in \mathbb{R}$, $x \in S$ and $\pi \in \Pi$ define

$$M^\lambda(x, g, \pi) := E_x^\pi \left[\exp \left\{ \lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g) \right\} \right], \tag{5.1}$$

where T is as in Assumption 2.2, and also define

$$M_+^\lambda(x, g) := \sup_{\pi \in \Pi} M^\lambda(x, g, \pi), \quad \text{and} \quad M_-^\lambda(x, g) := \inf_{\pi \in \Pi} M^\lambda(x, g, \pi). \tag{5.2}$$

There are two main objectives to be reached. The first one is to characterize the optimal value functions $M_+^\lambda(\cdot, \cdot)$ and $M_-^\lambda(\cdot, \cdot)$ via optimality equations, whereas the second goal is to establish the continuous dependence of $M_+^\lambda(\cdot, g)$ and $M_-^\lambda(\cdot, g)$ on the parameter g . These results are obtained below in Theorems 5.1 and 5.2, respectively, and these are used in Section 6 to construct solutions to the λ -ACOE given in Theorem 3.1. In particular, the continuity of the auxiliary value functions on the parameter g is used to obtain the λ -optimal average cost g_λ , as well as the function $h_\lambda(\cdot)$ in Theorem 3.1.

Theorem 5.1. *Let β be as in Definition 3.1, and suppose that λ is a real number satisfying*

$$\lambda \in \left(-\frac{\mu}{2\|C\|}, \frac{\mu}{2\|C\|} \right) \setminus \{0\}, \quad \text{where} \quad \mu = -\log(\beta). \tag{5.3}$$

In this case assertions (i)–(iii) below hold true.

(i) *For $x \in S$ and $g \in [-\|C\|, \|C\|]$,*

$$0 < \beta^K(1 - \beta e^{2|\lambda|\|C\|}) \leq M_-^\lambda(x, g) \leq M_+^\lambda(x, g) \leq \frac{1}{\beta^K(1 - \beta e^{2|\lambda|\|C\|})}.$$

(ii) *For each $g \in [-\|C\|, \|C\|]$, the optimal value functions $M_+^\lambda(\cdot, g)$ and $M_-^\lambda(\cdot, g)$ satisfy the following optimality equations:*

$$M_+^\lambda(x, g) = \sup_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right], \quad x \in S; \quad (5.4.a)$$

$$M_-^\lambda(x, g) = \inf_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_-^\lambda(y, g) \right) \right], \quad x \in S. \quad (5.4.b)$$

Moreover,

- (iii) Let $V : S \rightarrow \mathbb{R}$ be a bounded function and $g \in [-\|C\|, \|C\|]$. In this case,
 (a) If $V(\cdot)$ satisfies

$$V(x) = \sup_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) V(y) \right) \right], \quad x \in S, \quad (5.5.a)$$

then $V(\cdot) = M_+^\lambda(\cdot, g)$. Similarly,
 (b) If $V(\cdot)$ satisfies

$$V(x) = \inf_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) V(y) \right) \right], \quad x \in S, \quad (5.5.b)$$

then $V(\cdot) = M_-^\lambda(\cdot, g)$.

Proof. (i) Notice that $|\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)| \leq 2|\lambda| \|C\| T$ whenever $g \in [-\|C\|, \|C\|]$, so that in this case the following inequalities hold for every $x \in S$ and $\pi \in \Pi$:

$$E_x^\pi [e^{-2|\lambda| \|C\| T}] \leq E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)}] \leq E_x^\pi [e^{2|\lambda| \|C\| T}]. \quad (5.6)$$

On the other hand, by (5.3), $2|\lambda| \|C\| < \mu = -\log(\beta)$, and Theorem 4.1 implies that

$$E_x^\pi [e^{2|\lambda| \|C\| T}] \leq \frac{1}{\beta^K (1 - \beta e^{2|\lambda| \|C\|})} < \infty$$

and then Jensen's inequality yields

$$E_x^\pi [e^{-2|\lambda| \|C\| T}] \geq (E_x^\pi [e^{2|\lambda| \|C\| T}])^{-1} \geq \beta^K (1 - \beta e^{2|\lambda| \|C\|}) > 0,$$

and part (i) follows combining the last two sets of inequalities with (5.6), (5.1) and (5.2).

- (ii) To establish (5.4.a) let $g \in [-\|C\|, \|C\|]$ and notice that

$$\begin{aligned}
\sum_{t=0}^{T-1} (C(X_t, A_t) - g) &= \sum_{t=0}^{\infty} (C(X_t, A_t) - g) \mathcal{I}[T > t] \\
&= (C(X_0, A_0) - g) + \sum_{t=1}^{\infty} (C(X_t, A_t) - g) \mathcal{I}[T > t] \\
&= (C(X_0, A_0) - g) \\
&\quad + \sum_{t=1}^{\infty} (C(X_t, A_t) - g) \mathcal{I}[X_1 \neq z, \dots, X_t \neq z]. \quad (5.7)
\end{aligned}$$

Next, observe that for every policy π , $x, y \in S$ and $a \in A$, (a) and (b) below hold:

- (a) On the event $[X_1 = z]$,

$$\sum_{t=0}^{T-1} (C(X_t, A_t) - g) = C(X_0, A_0) - g,$$

so that

$$E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)} \mathcal{I}[X_1 = z] \mid A_0 = a, X_1 = y] = e^{\lambda(C(x, a) - g)} \delta_{yz}; \quad (5.8)$$

- (b) On the other hand, (5.7) yields that on the event $[X_1 \neq z]$,

$$\begin{aligned}
e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)} &= e^{\lambda(C(X_0, A_0) - g)} \\
&\quad \cdot e^{\lambda(C(X_1, A_1) - g) + \lambda \sum_{t=2}^{\infty} (C(X_t, A_t) - g) \mathcal{I}[X_2 \neq z, \dots, X_t \neq z]},
\end{aligned}$$

and then

$$\begin{aligned}
E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)} \mathcal{I}[X_1 \neq z] \mid A_0 = a, X_1 = y] \\
&= e^{\lambda(C(x, a) - g)} (1 - \delta_{yz}) \\
&\quad E_x^\pi [e^{\lambda(C(X_1, A_1) - g) + \lambda \sum_{t=2}^{\infty} (C(X_t, A_t) - g) \mathcal{I}[X_2 \neq z, \dots, X_t \neq z]} \mid A_0 = a, X_1 = y] \\
&= e^{\lambda(C(x, a) - g)} (1 - \delta_{yz}) E_y^{\pi'} [e^{\lambda(C(X_0, A_0) - g) + \lambda \sum_{t=1}^{\infty} (C(X_t, A_t) - g) \mathcal{I}[X_1 \neq z, \dots, X_t \neq z]}] \\
&= e^{\lambda(C(x, a) - g)} (1 - \delta_{yz}) E_y^{\pi'} [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)}]
\end{aligned}$$

where π' is the shifted policy defined by $\pi'_t(\cdot \mid h_t) = \pi_{t+1}(\cdot \mid x, a, h_t)$; the Markov property was used to obtain the second equality, and the third one comes from (5.7). Thus, using (5.1) and (5.2) it follows that

$$\begin{aligned} E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)} \mathcal{I}[X_1 \neq z] \mid A_0 = a, X_1 = y] \\ \leq e^{\lambda(C(x, a) - g)} (1 - \delta_{yz}) M_+^\lambda(y, g). \end{aligned}$$

- Combining this inequality with (5.8) it follows that

$$\begin{aligned} E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)} \mid A_0 = a, X_1 = y] \\ \leq e^{\lambda(C(x, a) - g)} (\delta_{yz} + (1 - \delta_{yz}) M_+^\lambda(y, g)) \end{aligned}$$

and, as a consequence,

$$\begin{aligned} E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)} \mid X_0, A_0 = a] \\ \leq e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \\ \leq \sup_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right] \end{aligned}$$

and taking expectations with respect to $P_x^\pi[\cdot]$, this implies that

$$\begin{aligned} M^\lambda(x, g, \pi) = E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)}] \\ \leq \sup_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right] \end{aligned}$$

and since $\pi \in \Pi$ and $x \in S$ were arbitrary, it follows that

$$M_+^\lambda(x, g) \leq \sup_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right], \quad x \in S. \quad (5.9)$$

To establish the reverse inequality, let $\varepsilon > 0$ and $f \in \mathbb{F}$ be arbitrary. For each $x \in S$ select a policy $\pi^x \in \Pi$ such that $M^\lambda(x, \pi^x, g) \geq M_+^\lambda(x, g) - \varepsilon$, and define a policy π by $\pi_0(\{f(x_0)\} \mid x_0) = 1$ and $\pi_t(\cdot \mid h_t) = \pi_{t-1}^{x_1}(\cdot \mid x_1, a_1, \dots, x_t)$, so that π chooses actions according to f at time $t = 0$, and if $X_1 = y$, π selects actions to apply from time $t = 1$ onwards according to π^y as if the process had started over again. In this case, the Markov property, (5.2) and (5.7) together imply that

$$\begin{aligned} M_+^\lambda(x, g) &\geq M^\lambda(x, g, \pi) \\ &= E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g)}] \end{aligned}$$

$$\begin{aligned}
 &= e^{\lambda(C(x,f(x))-g)} \left(p_{xz}(f(x)) + \sum_{y \neq z} p_{xy}(f(x)) M_+^\lambda(x, \pi^x, g) \right) \\
 &\geq e^{\lambda(C(x,f(x))-g)} \left(p_{xz}(f(x)) + \sum_{y \neq z} p_{xy}(f(x)) (M_+^\lambda(y, g) - \varepsilon) \right) \\
 &\geq e^{\lambda(C(x,f(x))-g)} \left(p_{xz}(f(x)) + \sum_{y \neq z} p_{xy}(f(x)) M_+^\lambda(y, g) \right) - \varepsilon e^{2|\lambda||C||}
 \end{aligned}$$

and then, since $\varepsilon > 0$ and policy $f \in \mathbb{F}$ are arbitrary, it follows that

$$M_+^\lambda(x, g) \geq \sup_{a \in A} \left[e^{\lambda(C(x,a)-g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right], \quad x \in S,$$

and this inequality together with (5.9) yields (5.4.a). The proof of (5.4.b) is similar.

(iii) (a) Let $V \in \mathcal{M}_b(S)$ be as in (5.4.a). In this case, (5.4) and Lemma 3.3 in [10] together imply that for every $x \in S$

$$\begin{aligned}
 |M_+^\lambda(x, g) - V(x)| &\leq \sup_{a \in A} \left[e^{\lambda(C(x,a)-g)} \sum_{y \neq z} p_{xy}(a) |M_+^\lambda(y, g) - V(y)| \right] \\
 &\leq e^{2|\lambda||C||} \sup_{a \in A} \left[\sum_{y \neq z} p_{xy}(a) |M_+^\lambda(y, g) - V(y)| \right] \quad (5.10)
 \end{aligned}$$

Next, for a given $\varepsilon > 0$, select $f \in \mathbb{F}$ such that for each $x \in S$

$$\begin{aligned}
 &e^{2|\lambda||C||} \sup_{a \in A} \left[\sum_{y \neq z} p_{xy}(a) |M_+^\lambda(y, g) - V(y)| \right] \\
 &\leq \varepsilon + e^{2|\lambda||C||} \sum_{y \neq z} p_{xy}(f(x)) |M_+^\lambda(y, g) - V(y)| \\
 &= \varepsilon + e^{2|\lambda||C||} E_x^f [|M_+^\lambda(X_1, g) - V(X_1)| \mathcal{I}[T > 1]],
 \end{aligned}$$

inequality that together with (5.10) yields

$$|M_+^\lambda(x, g) - V(x)| \leq \varepsilon + e^{2|\lambda||C||} E_x^f [|M_+^\lambda(X_1, g) - V(X_1)| \mathcal{I}[T > 1]],$$

and via an induction argument this leads to

$$\begin{aligned}
 |M_+^\lambda(x, g) - V(x)| &\leq E_x^f \left[\varepsilon \sum_{t=0}^n e^{2|\lambda||C||t} \mathcal{I}[T > t] \right. \\
 &\quad \left. + e^{2|\lambda||C||(n+1)} |M_+^\lambda(X_{n+1}, g) - V(X_{n+1})| \mathcal{I}[T > n+1] \right].
 \end{aligned}$$

To conclude, observe that an application of Lemma 4.2 produces

$$|M_+^\lambda(x, g) - V(x)| < \varepsilon \sum_{t=0}^n e^{2|\lambda|\|C\|^t} \beta^{t+1-K} + e^{2|\lambda|\|C\|(n+1)} \beta^{n+2-K} \|M_+^\lambda(\cdot, g) - V(\cdot)\|;$$

since $\beta e^{2|\lambda|\|C\|} < 1$ and $\|M_+^\lambda(\cdot, g) - V(\cdot)\| < \infty$, it follows, after taking limit as $n \rightarrow \infty$ in the right-hand side of above inequality, that

$$0 \leq |M_+^\lambda(x, g) - V(x)| < \frac{\varepsilon}{\beta^K(1 - \beta e^{2|\lambda|\|C\|})}$$

and since $x \in S$ and $\varepsilon > 0$ were arbitrary, it follows that $M_+^\lambda(\cdot, g) = V(\cdot)$. Part (b) is proved similarly. \square

The next theorem establishes that the optimal value functions in (5.2) depend continuously on the parameter g .

Theorem 5.2. *Let λ be as in (5.3). In this case,*

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

$$g \mapsto M_+^\lambda(x, g), \quad \text{and} \quad g \mapsto M_-^\lambda(x, g)$$

are continuous in $g \in [-\|C\|, \|C\|]$.

(ii) *There exist g_λ^+ and $g_\lambda^- \in \mathbb{R}$ such that*

$$M_+^\lambda(z, g_\lambda^+) = 1 = M_-^\lambda(z, g_\lambda^-). \tag{5.11}$$

Proof. (i) For $g, g_1 \in [-\|C\|, \|C\|]$ define

$$D_+^\lambda(x, g, g_1) = M_+^\lambda(x, g) - M_+^\lambda(x, g_1), \quad x \in S.$$

With this notation, Theorem 5.1(ii) implies that for every $x \in S$

$$\begin{aligned} M_+^\lambda(x, g) &= \sup_{a \in A} \left[e^{\lambda(C(x, a) - g)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right] \\ &= e^{\lambda(g_1 - g)} \sup_{a \in A} \left[e^{\lambda(C(x, a) - g_1)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g) \right) \right] \\ &= e^{\lambda(g_1 - g)} \sup_{a \in A} \left[e^{\lambda(C(x, a) - g_1)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g_1) \right) \right. \\ &\quad \left. + \sum_{y \neq z} p_{xy}(a) D_+^\lambda(y, g, g_1) \right] \end{aligned}$$

$$\begin{aligned} &\leq e^{\lambda(g_1-g)} \sup_{a \in A} \left[e^{\lambda(C(x,a)-g_1)} \left(p_{xz}(a) + \sum_{y \neq z} p_{xy}(a) M_+^\lambda(y, g_1) \right) \right] \\ &\quad + e^{\lambda(g_1-g)} \sup_{a \in A} \left[e^{\lambda(C(x,a)-g_1)} \sum_{y \neq z} p_{xy}(a) D_+^\lambda(y, g, g_1) \right] \\ &= e^{\lambda(g_1-g)} M_+^\lambda(x, g_1) + \sup_{a \in A} \left[e^{\lambda(C(x,a)-g)} \sum_{y \neq z} p_{xy}(a) D_+^\lambda(y, g, g_1) \right] \end{aligned}$$

so that for every $x \in S$

$$\begin{aligned} D_+^\lambda(x, g, g_1) &= M_+^\lambda(x, g) - M_+^\lambda(x, g_1) \\ &\leq [e^{\lambda(g_1-g)} - 1] M_+^\lambda(x, g_1) \\ &\quad + \sup_{a \in A} \left[e^{\lambda(C(x,a)-g_1)} \sum_{y \neq z} p_{xy}(a) D_+^\lambda(y, g, g_1) \right] \\ &\leq [e^{|\lambda(g_1-g)|} - 1] B + e^{2|\lambda||C|} \sup_{a \in A} \left[\sum_{y \neq z} p_{xy}(a) |D_+^\lambda(y, g, g_1)| \right] \end{aligned}$$

where $B := [\beta^K(1 - \beta e^{2|\lambda||C|})]^{-1}$ (see Theorem 5.1(i)) and the second inequality follows since g belongs to the interval $[-\|C\|, \|C\|]$. Interchanging g and g_1 it follows that

$$\begin{aligned} |D_+^\lambda(x, g, g_1)| &= |M_+^\lambda(x, g) - M_+^\lambda(x, g_1)| \\ &\leq [e^{|\lambda(g_1-g)|} - 1] B + e^{2|\lambda||C|} \sup_{a \in A} \left[\sum_{y \neq z} p_{xy}(a) |D_+^\lambda(y, g, g_1)| \right] \end{aligned}$$

which is equivalent to

$$\begin{aligned} |D_+^\lambda(x, g, g_1)| &\leq \sup_{\pi \in \Pi} E_x^\pi [e^{|\lambda(g_1-g)|} - 1] B \mathcal{I}[T > 0] \\ &\quad + e^{2|\lambda||C|} |D_+^\lambda(X_1, g, g_1)| \mathcal{I}[T > 1], \quad x \in S, \end{aligned}$$

and an induction argument using the Markov property yields that for every $n \in \mathbb{N}$ and $x \in S$,

$$\begin{aligned} |D_+^\lambda(x, g, g_1)| &\leq \sup_{\pi \in \Pi} E_x^\pi \left[e^{|\lambda(g_1-g)|} - 1 \right] B \sum_{t=0}^n e^{2|\lambda||C|t} \mathcal{I}[T > t] \\ &\quad + e^{2|\lambda||C|(n+1)} |D_+^\lambda(X_{n+1}, g, g_1)| \mathcal{I}[T > n + 1]. \end{aligned}$$

Applying now Lemma 4.2 and recalling that $\beta < 1$ and $|D_+^\lambda(\cdot, g, g_1)| =$

$|M_+^\lambda(\cdot, g) - M_+^\lambda(\cdot, g_1)| \leq 2B$, it follows that

$$\begin{aligned} |D_+^\lambda(x, g, g_1)| &\leq [e^{|\lambda(g_1-g)|} - 1]B \sum_{t=0}^n e^{2|\lambda|\|C\|t} \beta^{t+1-K} \\ &\quad + 2Be^{2|\lambda|\|C\|(n+1)} \beta^{n+2-K}, \quad x \in S, \quad n \in \mathbb{N}, \end{aligned}$$

and then, since $\beta e^{2|\lambda|\|C\|} < 1$ (see (5.3)), it follows, letting $n \rightarrow \infty$ in the right-hand side of the above inequality, that for every state x

$$\begin{aligned} |D_+^\lambda(x, g, g_1)| &\leq [e^{|\lambda(g_1-g)|} - 1]B \sum_{t=0}^{\infty} e^{2|\lambda|\|C\|t} \beta^{t+1-K} \\ &= [e^{|\lambda(g_1-g)|} - 1]B \frac{\beta}{\beta^K(1 - \beta e^{2|\lambda|\|C\|})} \\ &\leq [e^{|\lambda(g_1-g)|} - 1]B^2 \end{aligned}$$

and then $D_+^\lambda(x, g, g_1) = M_+^\lambda(x, g) - M_+^\lambda(x, g_1) \rightarrow 0$ as $g \rightarrow g_1$. The continuity of $g \mapsto M_-(x, g)$ is established similarly.

(ii) First suppose that $\lambda > 0$. In this case, $\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - \|C\|) \leq 0$, and $\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) + \|C\|) \geq 0$, so that, from (5.1), $M^\lambda(z, \|C\|, \pi) \leq 1$ and $M^\lambda(z, -\|C\|, \pi) \geq 1$ for every $\pi \in \Pi$, and then

$$M_-^\lambda(z, \|C\|) \leq M_+^\lambda(z, \|C\|) \leq 1, \quad M_+^\lambda(z, -\|C\|) \geq M_-^\lambda(z, -\|C\|) \geq 1.$$

Using now part (i), the intermediate value theorem implies the existence of g_λ^+ and g_λ^- in $[-\|C\|, \|C\|]$ satisfying (5.11). The case $\lambda < 0$ is handled in a similar way. \square

6 Proof of Theorem 3.1

The technical preliminaries developed in the previous section will be now used to establish the result stated in Section 3. To begin with, notice that (5.11) combined with (5.4.a–b) yield that

$$M_+^\lambda(x, g_\lambda^+) = \sup_{a \in A} \left[e^{\lambda(C(x,a) - g_\lambda^+)} \sum_y p_{xy}(a) M_+^\lambda(y, g_\lambda^+) \right], \quad x \in S. \quad (6.1)$$

$$M_-^\lambda(x, g_\lambda^-) = \inf_{a \in A} \left[e^{\lambda(C(x,a) - g_\lambda^-)} \sum_y p_{xy}(a) M_-^\lambda(y, g_\lambda^-) \right], \quad x \in S. \quad (6.2)$$

Proof of Theorem 3.1. Define $h_\lambda : S \rightarrow \mathbb{R}$ and $g_\lambda \in \mathbb{R}$ as follows: For $x \in S$

$$h_\lambda(x) = \begin{cases} \log(M_-^\lambda(x, g_\lambda^-)), & \text{if } \lambda > 0, \\ \log(M_+^\lambda(x, g_\lambda^+)), & \text{if } \lambda < 0, \end{cases} \quad (6.3)$$

and

$$g_\lambda = \begin{cases} g_\lambda^-, & \text{if } \lambda > 0, \\ g_\lambda^+, & \text{if } \lambda < 0, \end{cases} \quad (6.4)$$

The pair $(g_\lambda, h_\lambda(\cdot))$ satisfies the desired conclusions:

- (i) h_λ is a bounded function: this follows from the definition of h_λ and Theorem 5.1(i).
(ii) Combining (6.1)–(6.4) it follows that

$$\text{If } \lambda < 0, \quad e^{\lambda g_\lambda + h_\lambda(x)} = \sup_{a \in A} \left[e^{\lambda C(x,a)} \sum_y p_{xy}(a) e^{h_\lambda(y)} \right], \quad x \in S, \quad (6.5)$$

and

$$\text{For } \lambda > 0, \quad e^{\lambda g_\lambda + h_\lambda(x)} = \inf_{a \in A} \left[e^{\lambda C(x,a)} \sum_y p_{xy}(a) e^{h_\lambda(y)} \right], \quad x \in S, \quad (6.6)$$

equalities that can be summarized by

$$\text{sign}(\lambda) e^{\lambda g_\lambda + h_\lambda(x)} = \inf_{a \in A} \left[\text{sign}(\lambda) e^{\lambda C(x,a)} \sum_y p_{xy}(a) e^{h_\lambda(y)} \right], \quad x \in S, \quad (6.7)$$

which is the λ -ACOE.

- (iii) First, it will be proved that g_λ in (6.7) is a lower bound for the λ -sensitive optimal average cost, i.e., that $g_\lambda \leq J^*(\lambda, x)$ for all $x \in S$. The argument is better presented in two cases:

Case 1: $\lambda > 0$.

In this situation (6.7) is equivalent to (6.6), so that for every $\pi \in \Pi$ and $x \in S$

$$e^{\lambda g_\lambda + h_\lambda(x)} \leq E_x^\pi [e^{\lambda C(X_0, A_0) + h_\lambda(X_1)}] \quad (6.8)$$

Then, by induction, it can be shown that

$$e^{\lambda(n+1)g_\lambda + h_\lambda(x)} \leq E_x^\pi [e^{\sum_{t=0}^n \lambda C(X_t, A_t) + h_\lambda(X_{n+1})}]. \quad (6.9)$$

To prove the above, note that (6.8) satisfies it for $n = 0$. Suppose that (6.9) holds for some non-negative integer m . Let $\pi \in \Pi$ and $x \in S$ be fixed and note that, by the Markov property,

$$\begin{aligned} E_x^\pi [e^{\lambda \sum_{t=0}^{m+1} C(X_t, A_t) + h_\lambda(X_{m+2})} \mid X_0, A_0, \dots, X_{m+1}] \\ = e^{\lambda \sum_{t=0}^m C(X_t, A_t)} \\ \times E_{X_{m+1}}^\pi [e^{\lambda C(X_{m+1}, A_{m+1}) + h_\lambda(X_{m+2})} \mid X_0, A_0, \dots, X_{m+1}] \end{aligned}$$

$$\begin{aligned}
&= e^{\lambda \sum_{t=0}^m C(X_t, A_t)} E_{X_{m+1}}^{\pi'} [e^{\lambda C(X_0, A_0) + h_\lambda(X_1)}] \\
&\geq e^{\lambda \sum_{t=0}^m C(X_t, A_t)} e^{\lambda g_\lambda + h_\lambda(X_{m+1})},
\end{aligned}$$

where π' denotes a shifted policy defined in the obvious way, and the inequality comes from (6.8). Therefore,

$$\begin{aligned}
&E_x^\pi [e^{\lambda \sum_{t=0}^{m+1} C(X_t, A_t) + h_\lambda(X_{m+2})}] \\
&\geq E_x^\pi [e^{\lambda \sum_{t=0}^m C(X_t, A_t) + h_\lambda(X_{m+1})}] e^{\lambda g_\lambda} \\
&\geq e^{\lambda(m+1)g_\lambda + h_\lambda(x)} e^{\lambda g_\lambda} \\
&= e^{\lambda(m+2)g_\lambda + h_\lambda(x)},
\end{aligned}$$

where the last inequality above comes from the induction hypothesis. Hence, from (6.9) it follows that

$$\begin{aligned}
\lambda(n+1)g_\lambda + h_\lambda(x) &\leq \log(E_x^\pi [e^{\lambda \sum_{t=0}^n C(X_t, A_t)}]) + \|h_\lambda\| \\
&= \lambda J_n(\lambda, x, \pi) + \|h_\lambda\|
\end{aligned} \tag{6.10}$$

(see (2.4)) and then

$$g_\lambda \leq J_n(\lambda, x, \pi)/(n+1) + (\|h_\lambda\| - h_\lambda(x))/(n+1),$$

and after taking limit superior as $n \nearrow \infty$, (2.5) yields $g_\lambda \leq J(\lambda, x, \pi)$, $x \in S$ and, since π is an arbitrary policy, $g_\lambda \leq J^*(\lambda, \cdot)$, by (2.6).

Case 2: $\lambda < 0$.

Now (6.7) is equivalent to (6.5), and (6.8)–(6.10) hold true when the inequalities are reversed and $\|h_\lambda\|$ is replaced by $-\|h_\lambda\|$. Then, from $\lambda(n+1)g_\lambda + h_\lambda(x) \geq \log(E_x^\pi [e^{\lambda \sum_{t=0}^n C(X_t, A_t)}]) - \|h_\lambda\| = \lambda J_n(\lambda, x, \pi) - \|h_\lambda\|$, it follows, since $\lambda < 0$, that

$$g_\lambda \leq J_n(\lambda, x, \pi)/(n+1) - (\|h_\lambda\| + h_\lambda(x))/(n+1),$$

so that $g_\lambda \leq J(\lambda, x, \pi)$ and then $g_\lambda \leq J^*(\lambda, \cdot)$, since $\pi \in \Pi$ and $x \in S$ are arbitrary.

To conclude the proof of part (iii), let $\delta > 0$ be arbitrary small but fixed. From (6.7) it follows that there exists a policy $f \in \mathbb{F}$ such that

$$\begin{aligned}
&\text{sign}(\lambda) e^{\lambda(C(x, f(x)) + \Phi(x))} \sum_y p_{xy}(f(x)) e^{h_\lambda(y)} \\
&= \inf_{a \in A} \left[\text{sign}(\lambda) e^{\lambda C(x, a)} \sum_y p_{xy}(a) e^{\lambda h_\lambda(y)} \right], \quad x \in S,
\end{aligned} \tag{6.11}$$

where the *discrepancy function* $\Phi(\cdot)$ is such that

$$\|\Phi(\cdot)\| \leq \delta.$$

In this case, (6.7) and (6.11) together yield that $e^{\lambda g_\lambda + h_\lambda(x)} = E_x^f[e^{\lambda(C(X_0, A_0) + \Phi(X_0)) + h_\lambda(X_1)}]$, and by similar arguments as in (6.9) it follows that

$$e^{\lambda(n+1)g_\lambda + h_\lambda(x)} = E_x^f[e^{\sum_{t=0}^n \lambda(C(X_t, A_t) + \Phi(X_t)) + h_\lambda(X_{n+1})}].$$

Hence, using the inequality above, we obtain

$$\begin{aligned} & -\lambda(n+1)\|\Phi\| - \|h_\lambda\| + \lambda \sum_{t=0}^n (C(X_t, A_t) \\ & \leq \sum_{t=0}^n \lambda(C(X_t, A_t) + \Phi(X_t)) + h_\lambda(X_{n+1}) \\ & \leq \sum_{t=0}^n \lambda(C(X_t, A_t) + \lambda(n+1)\|\Phi\| + \|h_\lambda\|), \end{aligned}$$

it then follows that

$$\begin{aligned} e^{\lambda J_n(\lambda, x, f) - \lambda(n+1)\|\Phi\| - \|h_\lambda\|} & \leq E_x^f[e^{\sum_{t=0}^n \lambda(C(X_t, A_t) + \Phi(X_t)) + h_\lambda(X_{n+1})}] = e^{\lambda(n+1)g_\lambda + h_\lambda(x)} \\ e^{\lambda J_n(\lambda, x, f) + \lambda(n+1)\|\Phi\| + \|h_\lambda\|} & \geq E_x^f[e^{\sum_{t=0}^n \lambda(C(X_t, A_t) + \Phi(X_t)) + h_\lambda(X_{n+1})}] = e^{\lambda(n+1)g_\lambda + h_\lambda(x)} \end{aligned}$$

and then

$$\left| g_\lambda - \frac{1}{n+1} J_n(\lambda, x, f) \right| \leq \left(\|\Phi\| + \frac{\|h_\lambda\| + |h_\lambda(x)|}{\lambda(n+1)} \right), \tag{6.12}$$

so that $|g_\lambda - J(\lambda, x, f)| \leq \|\Phi\| \leq \delta$, and then $J^*(\lambda, \cdot) \leq J(\lambda, x, f) \leq g_\lambda + \delta$, and since $\delta > 0$ is arbitrary, it follows that $J^*(\lambda, \cdot) \leq g_\lambda$. This completes the proof of part (iii) since, as already was established, $J^*(\lambda, \cdot) \geq g_\lambda$.

(iv) Let the pair $(g_\lambda, h_\lambda(\cdot))$ be a solution to the λ -ACOE, where $g_\lambda \in \mathbb{R}$ and $h_\lambda(\cdot) \in \mathcal{M}_b(S)$ satisfies $h_\lambda(z) = 0$. By the arguments in (iii), g_λ is the optimal λ -sensitive average cost at every state, so that g_λ is uniquely determined. Now define $V : S \rightarrow \mathbb{R}$ by

$$V(x) = e^{h_\lambda(x)}, \quad x \in S,$$

and notice that $V(\cdot) \in \mathcal{M}_b(S)$ and that $V(z) = 1$, so that with g replaced by g_λ the λ -ACOE is equivalent to (5.5.a) if $\lambda < 0$, and to (5.5.b) if $\lambda > 0$. Then, Theorem 5.1(iii) yields that for every $x \in S$,

(a) If $\lambda < 0$,

$$e^{h_\lambda(x)} = V(x) = M_+^\lambda(x, g_\lambda) = \sup_{\pi \in \Pi} E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g_\lambda)}], \tag{6.13.a}$$

whereas

(b) For $\lambda > 0$,

$$e^{h_\lambda(x)} = V(x) = M_-^\lambda(x, g_\lambda) = \inf_{\pi \in \Pi} E_x^\pi [e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g_\lambda)}], \quad (6.13.b)$$

and these equalities establish the uniqueness of $h_\lambda(\cdot)$ (see also (6.3)).

(v) Under Assumption 2.1, for each $x \in S$ the right-hand side of (6.7) is a continuous function of $a \in A$. Therefore, it has a minimizer $f^*(x) \in A$, and the corresponding policy $f^* \in \mathbb{F}$ satisfies (6.11) and, consequently, (6.12) with $\Phi(\cdot) = 0$, so that for every $x \in S$, $\lim_{n \rightarrow \infty} J_n(\lambda, x, f^*) / (n + 1) = g_\lambda$, and then $J(\lambda, \cdot, f^*) = g_\lambda = J^*(\lambda, \cdot)$, that is, f^* is λ -AO. \square

Remark 6.1. Notice that (6.13a–b) can be condensed into a single equation:

$$\text{sign}(\lambda) e^{h_\lambda(x)} = \inf_{\pi \in \Pi} E_x^\pi [\text{sign}(\lambda) e^{\lambda \sum_{t=0}^{T-1} (C(X_t, A_t) - g_\lambda)}],$$

so that, defining the relative cost per stage as $C(\cdot, \cdot) - g_\lambda$, then $h_\lambda(\cdot) / \lambda$ is interpreted as the infimum of the certain equivalents of the total relative cost up to the first visit to state z in a positive time.

7 Conclusion

This paper considered controlled Markov chains endowed with a risk-sensitive average cost optimality criterion as defined in (2.4)–(2.6). Under the strong version of the simultaneous Doeblin condition in Assumption 2.2, it was shown in Theorem 3.1 that for risk sensitivity coefficients λ sufficiently close to zero, the λ -sensitive average cost optimality equation admits a bounded solution rendering a constant optimal average cost, as well as an optimal stationary policy whenever standard continuity-compactness conditions are satisfied. Also, it was shown via Example 3.1 that the conclusions in Theorem 3.1 *cannot be extended to arbitrary values of λ in general.*

The proof techniques used in the paper are very much self-contained and employ only basic probabilistic and analysis principles. However, the approach used in the paper differs from the usual one used to study the risk-neutral average cost criterion, in that the proof of Theorem 3.1 was obtained via auxiliary problems with the expected total-cost criterion. Finally, the version of the simultaneous Doeblin condition in Assumption 2.2 is *very strong*, and trying to extend the results in this paper to a more general framework is a worthwhile pursuit. Research in this direction is currently in progress.

Appendix

Proof of Lemma 4.1. Supposed that Assumptions 2.0 and 2.2 hold true. For each $n \in \mathbb{N}$ and $x \in S$ define

$$M_n(x) = \sup_{\pi \in \Pi} E_x[T \wedge n], \quad (\text{A.1})$$

and notice that

$$M_n(x) \leq M_{n+1}(x) \leq n + 1. \quad (\text{A.2})$$

Now observe that for every positive integer n , (2.8) yields

$$T \wedge n = 1 + \sum_{t=1}^{n-1} \mathcal{I}[X_1 \neq z, X_2 \neq z, \dots, X_t \neq z], \quad (\text{A.3})$$

so that for every policy π , $a \in A$ and $x, y \in S$,

$$\begin{aligned} & E_x^\pi[T \wedge (n+1) | A_0 = a, X_1 = y] \\ &= 1 + E_x^\pi \left[\sum_{t=1}^n \mathcal{I}[X_1 \neq z, X_2 \neq z, \dots, X_t \neq z] \mid A_0 = a, X_1 = y \right] \\ &= 1 + E_x^\pi \left[\mathcal{I}[X_1 \neq z] \left(1 + \sum_{t=2}^n \mathcal{I}[X_2 \neq z, X_3 \neq z, \dots, X_t \neq z] \mid \right. \right. \\ &\quad \left. \left. A_0 = a, X_1 = y \right) \right] \\ &= 1 + (1 - \delta_{yz}) E_x^\pi \left[1 + \sum_{t=2}^n \mathcal{I}[X_2 \neq z, X_3 \neq z, \dots, X_t \neq z] \mid \right. \\ &\quad \left. A_0 = a, X_1 = y \right] \\ &= 1 + (1 - \delta_{yz}) E_y^{\pi'} \left[1 + \sum_{t=1}^{n-1} \mathcal{I}[X_1 \neq z, X_2 \neq z, \dots, X_t \neq z] \right] \\ &= (1 - \delta_{yz}) E_x^\pi[T \wedge n], \end{aligned}$$

where the ‘shifted’ policy π' is determined by $\pi'_i(\cdot | h_t) = \pi_{t+1}(\cdot | x, a, h_t)$, and the Markov property together with (A.3) were used to obtain the last two equalities. Thus, (A.1)–(A.3) and this relation together yield

$$\begin{aligned} E_x^\pi[T \wedge (n+1) | A_0 = a, X_1 = y] &= 1 + (1 - \delta_{yz}) E_y^{\pi'}[T \wedge n] \\ &\leq 1 + (1 - \delta_{yz}) M_n(y) \\ &\leq 1 + (1 - \delta_{yz}) M_{n+1}(y), \end{aligned}$$

and then

$$\begin{aligned} E_x^\pi[T \wedge (n+1) | A_0 = a] &\leq 1 + \sum_{y \neq z} p_{xy}(a) M_{n+1}(y) \\ &\leq \sup_{a \in A} \left[1 + \sum_{y \neq z} p_{xy}(a) M_{n+1}(y) \right], \end{aligned}$$

so that

$$E_x^\pi[T \wedge (n+1)] \leq \sup_{a \in A} \left[1 + \sum_{y \neq z} p_{xy}(a) M_{n+1}(y) \right],$$

and since policy π is arbitrary, it follows that

$$M_{n+1}(x) \leq \sup_{a \in A} \left[1 + \sum_{y \neq z} p_{xy}(a) M_{n+1}(y) \right]. \quad (A.4)$$

Now, given $\varepsilon > 0$, select a policy $f \in \mathbb{F}$ such that

$$\sup_{a \in A} \left[1 + \sum_{y \neq z} p_{xy}(a) M_{n+1}(y) \right] \leq 1 + \varepsilon + \sum_{y \neq z} p_{xy}(f(x)) M_{n+1}(y)$$

for every state x , and observe that (A.4) yields that

$$\begin{aligned} M_{n+1}(x) &\leq 1 + \varepsilon + \sum_{y \neq z} p_{xy}(f(x)) M_{n+1}(y) \\ &= 1 + \varepsilon + E_x^f[M_{n+1}(X_1) \mathcal{I}[T > 1]], \quad x \in S, \end{aligned}$$

which via an induction argument and (A.3) leads to

$$\begin{aligned} M_{n+1}(x) &\leq E_x^f \left[\sum_{t=0}^{k-1} (1 + \varepsilon) \mathcal{I}[T > t] + M_{n+1}(X_k) \mathcal{I}[T > k] \right] \\ &\leq E_x^f \left[\sum_{t=0}^{k-1} (1 + \varepsilon) \mathcal{I}[T > t] \right] + (n+1) P_x^f[T > k], \end{aligned}$$

where (A.1) was used to obtain the second inequality. Then, Assumption 2.2 implies, after taking limit as $k \uparrow \infty$ in the last term above, that

$$M_{n+1}(x) \leq E_x^f \left[\sum_{t=0}^{\infty} (1 + \varepsilon) \mathcal{I}[T > t] \right] = (1 + \varepsilon) E_x^f[T] \leq (1 + \varepsilon) K.$$

To conclude observe that, by the monotone convergence theorem, for every

policy π and initial state x ,

$$\begin{aligned} E_x^\pi[T] &= \lim_{n \rightarrow \infty} E_x^\pi[T \wedge (n+1)] \\ &\leq \lim_{n \rightarrow \infty} M_{n+1}(x) \\ &\leq (1 + \varepsilon)K \end{aligned}$$

and the result follows, since $\varepsilon > 0$ was arbitrary. \square

References

- [1] Arapostathis A, Borkar VS, Fernández-Gaucherand E, Gosh MK, Marcus SI (1993) Discrete-time controlled Markov processes with average cost criteria: a survey. *SIAM, Journal on Control and Optimization* 31:282–334
- [2] Borkar VK (1984) On minimum cost per unit of time control of Markov chains. *SIAM Journal on Control and Optimization* 21:965–984
- [3] Cavazos-Cadena R (1989) Necessary conditions for the optimality equation in average reward Markov decision processes. *Journal of Applied Mathematics and Optimization* 19:599–613
- [4] Cavazos-Cadena R (1988) Necessary and sufficient conditions for a bounded solution to the optimality equation in average reward Markov decision chains. *Systems & Control Letters* 10:71–78
- [5] Fernández-Gaucherand E, Marcus SI (1997) Risk-sensitive optimal control of hidden Markov models: Structural results. *IEEE Transactions on Automatic Control* 42:1418–1422
- [6] Fleming WH, Hernández-Hernández D (1997a) Risk sensitive control of finite state machines on an infinite horizon. *SIAM, Journal on Control and Optimization* 35:1970–1810
- [7] Fleming WH, Hernández-Hernández D (1997b) Risk sensitive control of finite state machines on an infinite horizon II, Preprint
- [8] Hernández-Hernández D, Marcus SI (1996) Risk sensitive control of Markov processes in countable state space. *Systems & Control Letters* 29:147–155
- [9] Hernández-Lerma O, Lasserre JB (1996) *Discrete-time Markov control processes*. Springer, New York
- [10] Hinderer K (1970) *Foundations of non-stationary dynamic programming with discrete time parameter*. Lecture Notes on Operations Research and Mathematical Systems, No. 33, Springer, New York
- [11] Hordjik A (1974) *Dynamic programming and potential theory*. Mathematical Centre Tract No. 51, Matematisch Centrum, Amsterdam
- [12] Howard RA, Matheson JE (1972) Risk-sensitive Markov decision processes. *Management Sciences* 18:356–369
- [13] Jacobson DH (1973) Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic Control* 18:124–131
- [14] Marcus SI, Fernández-Gaucherand E, Hernández-Hernández D, Coraluppi S, Fard P (1996) Risk sensitive Markov Decision processes. In Byrnes CI, Datta BN, Gilliam DS, Martin CF (eds.) *Systems & control in the twenty-first century*. Series: Progress in Systems and Control, Birkhäuser, pp. 263–279
- [15] James MR, Baras JS, Elliot RJ (1994) Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems. *IEEE Transactions on Automatic Control* 39:780–792
- [16] Loève M (1980) *Probability theory I*. Springer, New York
- [17] Puterman M (1994) *Markov decision processes*. Wiley, New York
- [18] Ross SM *Applied probability models with optimization applications*. Holden-Day, San Francisco

- [19] Runolfsson T (1994) The equivalence between infinite horizon control of stochastic systems with exponential-of-integral performance index and stochastic differential games. *IEEE Transactions on Automatic Control* 39:1551–1563
- [20] Thomas LC (1980) Connectedness conditions for denumerable state Markov decision processes. In: Hartley R, Thomas LC, White DJ (eds.) *Recent advances in Markov decision processes*, Academic press, New York
- [21] Whittle P (1990) *Risk-sensitive optimal control*. Wiley, New York
- [22] Brau A, Fernández-Gaucherand E (1997) Controlled Markov chains with risk-sensitive exponential average cost criterion. *Proc. 36th IEEE Conference on Decision and Control*, San Diego, CA, 2260–2264
- [23] Fernández-Gaucherand E, Marcus SI (1995) Non-standard optimality criteria for stochastic control problems. *Proc. 34th IEEE Conference on Decision and Control*, New Orleans, LA, 585–589