Chapter 22

RISK-SENSITIVE OPTIMAL CONTROL IN COMMUNICATING AVERAGE MARKOV DECISION CHAINS

Rolando Cavazos-Cadena
Departamento de Estadística y Cálculo
Universidad Autónoma Agraria Antonio Narro
Buenavista, Saltillo, COAH 25315
MÉXICO

Emmanuel Fernández-Gaucherand
Department of Electrical & Computer Engineering
University of Cincinnati
Cincinnati, OH 45221-0030
USA

emmanuel@uc.edu

Abstract This work concerns discrete-time Markov decision processes with denumerable state space and bounded costs per stage. The performance of a control policy is measured by a (long-run) risk-sensitive average cost criterion associated to a utility function with constant risk sensitivity coefficient $\lambda$, and the main objective of the paper is to study the existence of bounded solutions to the risk-sensitive average cost optimality equation for arbitrary values of $\lambda$. The main results are as follows: When the state space is finite, if the transition law is communicating, in the sense that under an arbitrary stationary policy transitions are possible...
between every pair of states, the optimality equation has a bounded solution for arbitrary non-null $\lambda$. However, when the state space is infinite and noncompact, the communication requirement and a strong form of the simultaneous Doeblin condition do not yield a bounded solution to the optimality equation if the risk-sensitivity coefficient has a sufficiently large absolute value. In general, these results are weaker than those in Section 3.2 below.

Keywords: Markov decision processes, Exponential utility function, Constant risk sensitivity, Constant average cost, Communication condition, Simultaneous Doeblin condition. Bounded solutions to the risk-sensitive optimality equation.

1. INTRODUCTION

This work considers discrete-time Markov decision processes (MDPs) with finite state and action spaces and bounded costs. Besides a standard continuity-compactness requirement, the main structural feature of the decision model is that, under the action of each stationary policy, every pair of states communicate (see Assumption 2.3 below). On the other hand, it is assumed that the decision maker grades two different random costs according to the expected value of an exponential utility function with (non-null) constant risk-sensitivity coefficient $\lambda$, and the performance index of a control policy $\pi$ is the average cost criterion. Within this context, the main purpose of the paper is to study the existence of bounded solutions to the risk-sensitive average cost optimality equation corresponding to a non-null value of $\lambda$ (i.e., the $\lambda$-ACOE) which, under the continuity-compactness conditions in Assumption 2.1, yields an optimal stationary policy with constant risk-sensitive average cost. Thus, we are concerned in this paper with fundamental theoretical issues. The reader is referred to a growing body of literature in the application of risk-sensitive models in operations research and engineering, e.g., Fernández-Gaucherand and Marcus (1997), Avila-Godoy et al. (1997), Avila-Godoy and Fernández-Gaucherand (1999), Avila-Godoy and Fernández-Gaucherand (2000), Shayman and Fernández-Gaucherand (1999).

The study of stochastic dynamical systems with risk-sensitive criteria can be traced back, at least, to the work of Howard and Matheson (1972), Jacobson (1973), and Jaquette (1973, 1974). Particularly, in Howard and Matheson (1972) the case of OPVs with perfect information and under Assumption 2.3 below and assuming appropriability of the transition matrix induced by each stationary policy, a solution to the $\lambda$-ACOE was obtained via the Perron-Frobenius theory of positive matrices for arbitrary $\lambda \neq 0$. Recently, there has been an increasing interest on MDP's endowed with risk-sensitive criteria (Cavazos-Cadena and Fernández-Gaucherand, 1998a, b, d, Fernández-Gaucherand and Marcus, 1997, Fleming and McGeer, 1995; Fleming and Hernández-Hernández, 1997b; Hernández-Hernández and Marcus, 1996; James et al., 1994; Marcus et al., 1996; Rungfissan, 1994, Whit-
Notation. Throughout the remainder \( \mathbb{R} \) and \( \mathbb{N} \) stand for the set of real numbers and non-negative integers, respectively, and \( a \wedge b = \max\{a, b\} \) for \( a, b \in \mathbb{N} \).

Given a subset \( E \subseteq \mathbb{R} \), the 

\[ \sup_{t \in E} \{ |f(t)| \} \]

is the supremum of \( E \subseteq \mathbb{R} \), where \( f \in \mathbb{R} \). On the other hand, given \( x \in \mathbb{R} \), \( y \in \mathbb{R} \), the 

\[ x \wedge y = \min\{x, y\} \]

is the minimum of \( x \) and \( y \). Finally, for an event \( W \), the corresponding indicator function is denoted by \( I[W] \).

2. THE DECISION MODEL

Following standard notation (Azaropoulou, 1993; Hernández-Lemus and Lasserre, 1996; Puterman, 1994), let an MDP be specified by the four-tuple \( M = (S, A, C, P) \), where the state space \( S \) is countable, the measurable state space \( A \) is the decision (action) set, \( C \subseteq S \times A \) is the cost function, and \( P = \{ p_{a}(s) \} \) is the Markov transition law.

The interpretation of the model \( M \) is as follows: At each time \( t \in \mathbb{N} \), the state of the dynamical system is observed, say \( x_{t} = x \in S \), and an action \( a_{t} = a \in \mathbb{N} \) 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 \( x_{t+1} = x' \in S \) with probability \( P_{a}(s) \).

Notice that it is assumed that every \( a \in \mathbb{N} \) is an admissible action at each state \( x \) and the model under this condition does not imply any loss of generality.

The following standard assumption is assumed to hold throughout the sequel.

Assumption 2.1. (i) The control set \( A \) is compact.

(ii) For each \( x_{t} \in S \) and \( a_{t} \in A \), \( C(x_{t}, a_{t}) \) is a continuous function in \( x_{t} \).

Policies. For each \( t \in \mathbb{N} \), the space of histories up to time \( t \) is recursively defined by \( H_{t} = (S_{t}, A_{t}) \). Each element of \( H_{t} \) is denoted by \( h_{t} = (x_{t}, a_{t}) \).

A policy is a sequence \( \pi = (\{a_{t}\}_{t=1}^{\infty}) \) where each \( a_{t} \) is a stochastic kernel on \( A \), given \( H_{t} \), so that, for each \( x_{t} \), \( a_{t}(x_{t}) \) is a measurable function on \( A \), and for each Borel subset \( B \subset A \), \( H_{t} \rightarrow a_{t}(B|x_{t}) \) is a measurable function in \( x_{t} \) with \( H_{t} \). The number \( a_{t}(B|x_{t}) \) is the probability of choosing action \( a_{t} \in B \) when the system is driven by \( H_{t} \). Throughout the remainder \( P \) denotes the class of all policies.

The random variable \( X_{0} = x \), the distribution of the state-action process \( \{X_{t}, A_{t}\} \) is uniquely determined via Ionescu-Tulcea's theorem (Puterman, 1994; Hernández-Lemus and Lasserre, 1996; Hindriks, 1990; Puterman, 1994).
MODELING UNCERTAINTY

(see (2.3)) whereas the long-run λ-sensitive average cost under τ starting at x is defined by

\[ J(\lambda, \tau, x) = \lim_{n \to \infty} \frac{1}{n} J_n(\lambda, \tau, x). \]  

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

\[ J^*(\lambda, x) = \inf_{\tau} J(\lambda, \tau, x), \]

and a policy π* ∈ P is λ-average optimal (λ-AO) if J(λ, π*, x) = J^*(\lambda, x) for every x ∈ S.

Remark 2.1: From (2.3)-(2.6) it is not difficult to see that \( \min_{\sigma, d} C(x, \sigma) \leq J(\lambda, \pi, x) \leq \max_{\sigma, d} C(x, \sigma) \), so that the optimal average cost satisfies \( \min_{\sigma, d} C(x, \sigma) \leq J^*(\lambda, x) \leq \max_{\sigma, d} C(x, \sigma) \); see, for instance, Cavazos-Cadena and Fernández-Gaucherand (1998a).

Communication and Stability Assumptions. In the risk-neutral case (λ = 0), it is well-known that a strong recurrence condition is necessary for the existence of a bounded solution to the corresponding average cost optimality equation (Arapostathis et al., 1993; Bertsekas, 1987; Fernández-Gaucherand et al., 1990; Hernández-Lerma and Lasserre, 1996; Kumar and Varaiya, 1986; Puterman, 1994), which in turn yields an optimal stationary policy in a standard way. For the risk-sensitive average criterion in (2.4)-(2.6), it was shown in Cavazos-Cadena and Fernández-Gaucherand (1998a) that the following stability condition is necessary for the existence of a bounded solution to the λ-AOCE (see below in (3.1)).

Assumption 2.2: (Simultaneous Doeblin Condition (Arapostathis et al., 1993; Hernández-Lerma and Lasserre, 1996; Hordijk, 1974; Puterman, 1994; Ross, 1994; Thomas, 1980.) There exist a state x ∈ S and a positive constant K such that

\[ E^f[x_1 \mid T^f] \leq K. \]

for all x ∈ S and f ∈ F, \( x' = S \) where for each \( x' \in S \), the first passage time \( T^f \) to state y is defined by

\[ T^f = \min\{n \geq 0 \mid X_n = y\}, \]

with the (usual) convention that the minimum of the empty set is \( \infty \).

It was shown in Cavazos-Cadena and Fernández-Gaucherand (1998a-d) that Assumptions 2.1 and 2.2 yield a bounded solution to the λ-AOCE in (3.1) below whenever the risk sensitivity coefficient is small enough. To study the existence of bounded solutions to the λ-AOCE for arbitrary λ ≠ 0, the following additional condition, used firstly in Howard and Matheson (1972), will be employed.

Assumption 2.3. (Communication.) Under every stationary policy each pair of states communicate, i.e., given f ∈ F and x, y ∈ S, there exists n = n(x, y, f) ∈ N such that \( P^f[X_n = y] > 0 \).

Remark 2.2. It will be shown in the sequel that, when the state and action spaces are finite, Assumption 2.3 implies that Assumption 2.2 holds true.

Relation to the work of Howard-Matheson. Under Assumption 2.4 below, it was proved in Howard and Matheson (1972), via the Perron–Frobenius theory of positive matrices, that for arbitrary λ > 0, the λ-AOCE has a (bounded) solution (rewards instead of costs were considered in Howard and Matheson (1972)).

Assumption 2.4. (a) The state and action spaces are finite (notice that in this situation Assumption 2.1 is automatically satisfied); (b) Assumption 2.3 holds and the transition matrix induced by an arbitrary stationary policy is aperiodic.

On the other hand, it has been recently shown in Cavazos-Cadena and Fernández-Gaucherand (1998a) that, even when the state space is finite, under Assumptions 2.1 and 2.2, the λ-AOCE has a bounded solution only if the risk sensitivity coefficient is small enough, and an example was given showing that this conclusion cannot be extended to arbitrary values of λ. The difference between the conclusions in Cavazos-Cadena and Fernández-Gaucherand (1998a) and Howard and Matheson (1972), comes from the different settings in both papers. In particular, Assumption 2.3 is imposed in Howard and Matheson (1972), but not in Cavazos-Cadena and Fernández-Gaucherand (1998a), and an additional aperiodicity condition is used in the latter reference.

3. MAIN RESULTS

The main problems considered in the paper consists in studying if Assumptions 2.1-2.3 yield (bounded) solutions to the λ-AOCE for arbitrary values of the risk sensitivity coefficient λ ≠ 0. It turns out that the answer depends on the state space: If S is finite, Assumption 2.3, combined with Assumption 7.1, implies the existence of a solution to the λ-AOCE for arbitrary λ ≠ 0; this result is presented below as Theorem 3.1 and gives an extension of that in Howard and Matheson (1972). On the other hand, as it will be shown via a detailed example, such a conclusion cannot be extended to the case in which S is countably infinite, thus providing an extension to the results in Cavazos-Cadena and Fernández-Gaucherand (1998a).

- Finite State Space Models.
The following theorem shows that Assumption 2.1 and Assumption 2.3 are sufficient to guarantee a (bounded) solution to the λ-ACOE for arbitrary values of the non-trivial risk sensitivity coefficient λ. Hence, our results extend those in Howard and Matheson (1972) in that theapperiodicity in Assumption 2.4(b) is not required, and in that our results hold for both the risk-seeking and risk-averse cases.

**Theorem 3.1.** Let the state space $S$ be finite and suppose that Assumptions 2.1 and 2.3 hold true. In this case, for every $\lambda \neq 0$ there exist a constant $g_0 \in \mathbb{R}$ and a function $h_0: S \rightarrow \mathbb{R}$ such that the following are true.

(i) The pair $(g_0, h_0)$ satisfies the λ-ACOE:

$$\sgn(\lambda) e^{-h_0(x)} + h_0(x) = \inf_{\delta \in \mathcal{A}} \left\{ \sgn(\lambda) e^{\delta(x)} \sum_{a} \rho_{\delta}(a) e^{h_0(a)} \right\}, \quad x \in S.$$  

(ii) $J^*(\lambda, x) = g_0$ for each $x \in S$.

(iii) For every $x \in S$, the term in brackets in the right-hand side of (3.1) is a continuous function on $A$; thus, it has a minimizer $f(x) \in A$ and the corresponding policy $f \in \mathcal{F}$ is λ-AO. Moreover,

(iv) The pair $(g_0, h_0)$ is unique whenever $h_0(\cdot)$ satisfies $h_0(x) = 0$, where $x \in S$ is arbitrary but fixed.

Remark 3.1. Theorem 3.1 provides an extension to the results in Howard and Matheson (1972), in that both the risk-averse and risk-seeking cases are considered, the action space is not restricted to be finite, and the requirement ofaperiodicity of the transition matrices associated to stationary policies is avoided; notice that this latter condition is essential in the development of the Perron-Frobenius theory of positive matrices, which was the key tool employed in Howard and Matheson (1972). Also, observing that Assumption 2.3 yields the Doobian condition in Assumption 2.2 when the state space is finite (see Theorem 4.1 in the next section), Theorem 3.1 above can be seen as an extension to Theorem 3.1 in Cavazos-Cadena and Fernández-Gauchard (1998a), in that by restricting the framework to a finite state space, Assumptions 2.1 and 2.3 yield a solution to the λ-ACOE for every $\lambda \neq 0$. In Cavazos-Cadena and Fernández-Gauchard (1998a), it was shown that the λ-ACOE admits bounded solution only for $|\lambda| \neq 0$ small enough, in contrast to the claims in Hernández-Hernández and Marcus (1996).

The somewhat technical proof of Theorem 3.1 will be presented in Section 7, after the necessary preliminaries stated in the following three sections. As in Cavazos-Cadena and Fernández-Gauchard (1998a), the main idea is to consider a family of auxiliary parameterized stopping problems for the MDP endowed with a risk-sensitive expected-total cost criterion, and then Theorem 3.1 will be obtained by an appropriate selection of the parameter.

• Denumerable State Space Models. In view of Theorem 3.1 above it is natural to ask the following: Do, for every $\lambda \neq 0$, Assumptions 2.1–2.3 together yield a bounded solution to the λ-ACOE when the state space is countably infinite? Such a solution is indeed obtained under the above assumptions for the risk-neutral case (Arapostathis et al., 1993). The following example shows that the answer to the question above is negative.

**Example 3.1.** For each positive integer $r$ define

$$b_r = K_0 r^{-\gamma} \quad \text{and} \quad B_r = K_0 \sum_{m=1}^{\infty} m^{-\gamma} r^{-\gamma}.$$  

(3.2)

where $K_0$ is selected in such a way that $B_1 = 1$. Now, for each $\lambda \neq 0$ define an MDP $M = (S, A, \rho_\delta, C)$ as follows. The state space is $S = \mathbb{N}$, the action space $A = \{a\}$ is a singleton, the transition law is defined by

$$b_{n+1}(a) = \frac{B_{n+2}}{B_{n+1}}, \quad \rho_\delta(a) = \frac{b_{n+1}}{B_{n+1}}; \quad n \in \mathbb{N},$$

whereas the cost function $C((\cdot \cdot))$ is given by

$$C((n, a)) = C(n, a) = sgn(\lambda), \quad n \neq 0,$$  

$$C((0, a)) = 0.$$  

(3.3)

In the following proposition it will be proved that Assumptions 2.1–2.3 hold in this example, and then, that the λ-ACOE does not have a bounded solution for $|\lambda|$ large enough.

**Proposition 3.1.** For the MDP in Example 3.1 above, Assumptions 2.1–2.3 hold true and, moreover, the transition matrix in (3.2) is aperiodic.

**Proof.** Assumption 2.1 clearly holds in this example, since $A$ is a singleton. To verify Assumption 2.2, let $x = 0$ and notice that, since from state $n \in S$, transitions are possible to $n + 1$ or to state $x = 0$, it is not difficult to see that

$$P_n[\tau \geq t] = P_n[x = n + t] = B_{n+1}/B_{n+1}, \quad n, t \in \mathbb{N}$$

and from (3.2) it follows that

$$\frac{B_{n+2}}{B_{n+1}} = \frac{\sum_{m=n+1}^{\infty} (m/n)^{1-\gamma} m^{1-\gamma}}{\sum_{m=n+1}^{\infty} m^{1-\gamma} / m^{1-\gamma} / 2}.$$  

(3.4)
MODELING UNCERTAINTY

\[ \sum_{t=0}^{\infty} e^{-\gamma t} \frac{e^{-\tau t}}{\tau + 1} \leq \sum_{t=0}^{\infty} e^{-\gamma t} \frac{e^{-\tau t}}{\tau + 1} = e^{-\gamma} \sum_{t=0}^{\infty} e^{-\tau t} = e^{-\gamma} \]

so that for every \( n \in S \),

\[ E_n(T_\lambda) = \sum_{t=0}^{\infty} e^{-\gamma t} P_\lambda[T_\lambda > t] \leq \sum_{t=0}^{\infty} e^{-\gamma t} = e^{-\gamma} - 1. \]

verifying Assumption 2.2. Now observe that by (3.2) and (3.4), \( P_\lambda[X_1 = 0] = b_0 + \frac{b_1}{B_1} > 0 \), and \( P_\lambda[X_1 = m] = b_m + \frac{b_{m+1}}{B_1} > 0 \), so that \( P_\lambda[X_{m+1} = m] = P_\lambda[X_1 = 0] \times P_\lambda[X_1 = m] > 0 \), and then the communication condition in Assumption 2.3 holds. Notice, finally, that \( P_\lambda[X_1 = 0] = b_1 / B_1 > 0 \), so that the transition law (3.2) is aperiodic.

Observe now that, for Example 3.1, the \( \lambda \)-ACOE reduces to the following Poisson equation:

\[ e_{\lambda}(x) = E_{\lambda} \left[ e^{\lambda t} | X_0 = x \right], \quad x \in S \]  

(3.5)

In the following proposition it will be shown that this equation does not admit a bounded solution if \( |\lambda| \) is large enough. First, let \( \lambda_0 \) be determined by

\[ e_{\lambda_0} = K_0 \sum_{i=1}^{\infty} \frac{1}{i^2}. \]  

(3.6)

Proposition 3.2. For the MDP in Example 3.1, with \( z = 0 \), the following assertions hold:

(i) \( E_{\lambda} e^{\lambda t} < \infty \iff |\lambda| \leq 1 \);

(ii) \( E_{\lambda} e^{\lambda t} < \infty \Rightarrow E_{\lambda} e^{\lambda t} < e_{\lambda_0} \);

(iii) For \( |\lambda| > \lambda_0 \), there is not a pair \( (q, h(t)) \in R \times B(S) \) satisfying (3.5).

Proof. (i) From (3.2) and (3.4), it is not difficult to see that \( P_\lambda[T_\lambda = t] = b_t / B_1 = b_t \) for every positive integer \( t \), so that

\[ E_{\lambda} e^{\lambda t} = \sum_{t=1}^{\infty} e^{\lambda t} b_t = K_0 \sum_{t=1}^{\infty} e^{\lambda t} / t^2, \]  

(3.7)

and it is clear that \( E_{\lambda} e^{\lambda t} < \infty \) is equivalent to \( |\lambda| \leq 1 \).
The following inequality holds:
\[ e^{-|\delta^+| |\alpha|^{(2r-\|\alpha\|)+K^+(T_{r+1})}|1-\rho^+|^{2r-\|\alpha\|}} \leq e^{-|\delta^+| |\alpha|^{(2r-\|\alpha\|)+K^+(T_{r+1})}|1-\rho^+|^{2r-\|\alpha\|}}. \]

Using (a)-(b), (39) implies, via the dominated convergence theorem, that
\[ e^{N(k)} = E_{x} \left[ e^{\delta^+T_{r+1}[(1-\rho^+)^{2r-\|\alpha\|}]} \right], \]

i.e.,
\[ e^{N(k)} = E_{x} \left[ e^{\delta^+T_{r+1}} \right] \]

In this case, part (ii) implies that \( e^{N(k)} \leq e^{N(k)} \), so that \( |\alpha| \leq N_0 \). In short, it has been proved that if a pair \((\rho, \alpha; 5) \in \mathbb{R} \times B(S)\) exists such that (3.5) is satisfied, then \( |\alpha| \leq N_0 \).

4. BASIC TECHNICAL PRELIMINARIES

This and the following two sections contain the ancillary technical results that will be used to establish Theorem 3.1. The main objective is to collect some basic consequences of Assumptions 2.1 and 2.3 which will be useful in the sequel. These results are presented below as Theorems 4.1 and 4.2 and Lemma 4.1. Throughout the remainder of the paper, the state space \( S \) is assumed to be finite.

The first result establishes that Assumptions 2.1 and 2.3 together imply a strong form of the simultaneous Doobian condition.

**Theorem 4.1.** Under Assumptions 2.1 and 2.3, there exists \( 0 \leq K < \infty \) such that
\[ E_{x}[T_y] \leq K, \quad \forall x, y \in S, \quad x \in P. \]

The proof of this theorem, based on ideas related to the risk–neutral average cost criterion, is contained in Appendix A. Suppose now that the initial state is \( X_0 = y \). The following result provides a lower bound for the probability of reaching a state \( x \neq y \) before returning to \( y \).

**Theorem 4.2.** Let \( x, y \in S \), with \( x \neq y \) arbitrary but fixed and suppose that Assumptions 2.1 and 2.3 hold true. In this case,

(i) There exists a constant \( \Lambda = \Lambda(x, y) > 0 \) such that, for every \( \pi \in \mathcal{P} \),
\[ P_{\pi}^{\Lambda} [T_x < T_y] \geq \Lambda. \]

(ii) Given \( \pi \in \mathcal{P} \), there exists a positive integer \( k \) such that
\[ P_{\pi}^{\Lambda} [X_k = x, X_{\leq k} \neq x, 1 \leq k < \ell] < P_{\pi}^{\Lambda} [k = T_x, T_y > 0]. \]

The arguments leading to the proof of this result are also based on ideas using the risk–neutral average cost criterion, and are presented in Appendix B.

The following lemma will be useful in the proof of Theorem 3.1 (which is presented in Section 7).

**Lemma 4.1.** Suppose that Assumptions 2.1 and 2.3 hold true, and let \( D \in B(S) \) be such that for some \( f \in F \) and \( h \in B(S) \)
\[ e^{N(k)} = E_{x} \left[ h(X_0 + d(X_1)) \right], \quad x \in S. \]

(i) For each positive integer \( n \)
\[ e^{N(k)} = E_{x} \left[ e^{\sum_{i=0}^{n-1} D(X_i)} h(X_n + d(X_n)) \right], \quad x \in S. \]

(ii) For each \( x \in S \)
\[ E_{x} \left[ e^{\sum_{i=0}^{n-1} D(X_i)} I(T_x > n) \right] \to 0, \quad \text{as } n \to \infty, \]

and (iii)
\[ e^{N(k)} = E_{x} \left[ e^{\sum_{i=0}^{n-1} D(X_i)} h(X_n + d(X_n)) I(T_x \leq n) \right], \quad x \in S. \]

**Proof.** (i) Define the sequence of random variables \( \{ \mathcal{M}_n \} \) by \( \mathcal{M}_0 = e^{N(k)} \) and \( \mathcal{M}_n = e^{\sum_{i=0}^{n-1} D(X_i + d(X_i))} \) for \( n = 1, 2, \ldots \). In this case, for each \( n \in \mathbb{N} \) and \( x, y \in S \), the Markov property and (41) together yield
\[ E_{x} \left[ M_{n+1} | X_0, \ldots, X_n = y \right] = e^{\sum_{i=0}^{n-1} D(X_i) I(T_x > n)} E_{x} \left[ M_n | X_0, \ldots, X_n = y \right] = e^{\sum_{i=0}^{n-1} D(X_i) I(T_x > n) + h(X_n + d(X_n))}, \]

so that
\[ E_{x} \left[ M_{n+1} | X_0, \ldots, X_n = y \right] = M_n. \]
i.e., \( \{ M_n \} \) is a martingale with respect to the probability measure \( P \) and the standard filtration \( \mathcal{F}_t \) of the initial state. Therefore, by optional stopping, for every \( x \in S \) and \( n \in \mathbb{N} \setminus \{ 0 \} \), \( \mathbb{E}^{0,x} \{ M_n \} = \mathbb{E}^{0,x} \{ M_{T_0} \} \) so that
\[
\rho^{(x)} = \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} \right].
\]

(ii) By Theorem 4.1, \( P(T_n < \infty) = 1 \) always holds, so that, since \( X_{T_n} = z \) on the event \( \{ T_n < \infty \} \) (see (2.3)),
\[
\lim_{n \to \infty} \rho^{(x)} = \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} \right] P \text{'-almost surely},
\]
and then (4.2) and Patlak's Lemma yield that
\[
\rho^{(x)} \geq \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} \right].
\]

Observe now that for a given \( x \in S \), Theorem 4.1 implies that there exists a positive integer \( k = k(x, z) \) such that \( P(T_k = k) > 0 \), so that
\[
\mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} \right] \geq \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_k = k \right] \geq e^{-4D \rho} \mathbb{E} \left[ P(T_k = k) > 0 \right],
\]
and since \( S \) is finite, there exist \( t \in (0, \infty) \) such that
\[
\mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} \right] \geq t.
\]

Next set \( \mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n) \) and observe that
\[
\mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | \mathcal{F}_n, X_0 = y \right] = \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_n > n \right] \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_n > n \right]
\]
so that
\[
\mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_n > n \right] \leq e^{-4D \rho} \mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_n > n \right] \to 0,
\]
by the dominated convergence theorem.

(iii) Since \( e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_n > n \) \( \leq \rho^{(x)} e^{\sum_{t=0}^{T \wedge \tau} b(X_t)} | T_n > n \), part (ii) yields that
\[
\mathbb{E} \left[ e^{\sum_{t=0}^{T \wedge \tau} (D(X_t)+b(X_t))} | T_n > n \right] \to 0, \quad \text{as } n \to \infty.
\]
Theorem 5.1. Let $x \in S$ be arbitrary but fixed. In this case,

(i) For each $y \in S$,
\[ 0 \leq M_\pi(x, y, D) \leq M_\pi(y, x, D). \]

(ii) If $M_\pi(x, y, D) < \infty$, then

(a) $M_\pi(y, x, D) < \infty$ for every $y \in S$.

(b) The following optimality equation holds:
\[ M_\pi(x, y, D) = \sup_{\alpha \in \Delta} \left[ e^{\alpha (y, D)} \left( p_{y, \alpha}(y) + \sum_{y' \in S} p_{y', \alpha}(y') M_\pi(y', x, D) \right) \right], \quad y \in S. \]

Similarly,

(iii) If $M_\pi(x, y, D) < \infty$, then

(a) $M_\pi(y, x, D) < \infty$ for every $y \in S$.

(b) The following optimality equation holds:
\[ M_\pi(x, y, D) = \inf_{\alpha \in \Delta} \left[ e^{\alpha (y, D)} \left( p_{y, \alpha}(y) + \sum_{y' \in S} p_{y', \alpha}(y') M_\pi(y', x, D) \right) \right], \quad y \in S. \]

Proof. (i) Let $\pi$ be an arbitrary policy, and observe that for every $x, y \in S$,

Jensen’s inequality yields
\[
E_x \left[ \sum_{i=0}^{\infty} \gamma^i \delta(x_{i+1}, y) \right] \geq E_x \left[ \sum_{i=0}^{\infty} \gamma^i \delta(x_{i+1}, y_{i+1}) \right] \geq e^{-\gamma^{\infty}} E_x \left[ \sum_{i=0}^{\infty} \gamma^i \delta(x_{i+1}, y_{i+1}) \right] \geq e^{-\gamma^{\infty}} E_x \left[ \sum_{i=0}^{\infty} \gamma^i \delta(x_{i+1}, y_{i+1}) \right],
\]

where $K$ is as in Theorem 4.1. Therefore,
\[
M_\pi(x, y, D) = \inf_{\pi} E_x \left[ \sum_{i=0}^{\infty} \gamma^i \delta(x_{i+1}, y) \right] \geq e^{-\gamma^{\infty}} E_x \left[ \sum_{i=0}^{\infty} \gamma^i \delta(x_{i+1}, y) \right].
\]
which, since \( \pi \in \mathcal{P} \) was arbitrary, yields

\[
\sup_{x} M_{\pi}(x, z, D) \geq e^{-e(D)M_{\pi}} \mathcal{D}(y, z, D) P_{\pi}[X_{i} \neq x, 1 \leq i < k, X_{k} = y].
\]

Selecting \( k > 0 \) such that \( P_{\pi}[X_{i} \neq x, 1 \leq i < k, X_{k} = y] > 0 \) (which is possible, by Theorem 4.2(ii)), it follows that \( M_{\pi}(x, y, D) < \infty \), establishing part (a), whereas part (b) follows along the lines used in the proof of Theorem 3.1 in Cavazos-Cadena and Fernández-Gaucherand (1998a).

(iii) Suppose that \( M_{\pi}(x, z, D) < \infty \). In this case, there exists a policy \( \pi \in \mathcal{P} \) such that

\[
\sup_{x} M_{\pi}(x, z, D) \geq e^{-e(D)M_{\pi}} \sum_{i=1}^{k-1} D(X_{i}, A_{i}).
\]

Next, by the Markov property, for each positive \( k \) and \( y \in S \) \( \{ x \} \),

\[
E^{\pi}[I(X_{k} = y, X_{i} \neq x, 1 \leq i < k)] \sum_{i=0}^{k-1} D(X_{k}, A_{k}) P_{\pi}
\]

\[
\geq I(X_{k} = y, X_{i} \neq x, 1 \leq i < k, X_{k} = y) e^{-e(D)M_{\pi}} \sum_{i=0}^{k-1} D(X_{i}, A_{i})
\]

\[
\geq I(X_{k} = y, X_{k} \neq y, 1 \leq i < k, X_{k} = y) e^{-e(D)M_{\pi}} \sum_{i=0}^{k-1} D(X_{i}, A_{i})
\]

\[
\geq I(X_{k} = y, X_{i} \neq x, 1 \leq i < k, X_{k} = y) e^{-e(D)M_{\pi}} \sup_{x} M_{\pi}(x, y, D),
\]

where the shifted policy is defined by \( \pi^{*} \{ h_{k} \} = \pi^{*}(x, A_{0}, \ldots, A_{k-1}, h_{k}) \).

Taking the expectation with respect to \( P_{\pi}^{z} \), it follows that

\[
\sup_{x} M_{\pi}(x, z, D) \geq e^{-e(D)M_{\pi}} \sum_{i=0}^{k-1} D(X_{i}, A_{i})
\]

\[
\geq e^{-e(D)M_{\pi}} \sum_{i=0}^{k-1} D(X_{i}, A_{i})
\]

\[
\geq e^{-e(D)M_{\pi}} \sup_{x} M_{\pi}(x, z, D),
\]

and after selecting \( k \) such that \( P_{\pi}^{z} X_{k} \neq x, 1 \leq i < k, X_{k} = y > 0 \) (by Theorem 4.2(ii)), it follows that \( M_{\pi}(x, y, D) < \infty \), establishing part (a), whereas part (b) can be proved using the same arguments as in the proof of Theorem 3.1 in Cavazos-Cadena and Fernández-Gaucherand (1998a).

Theorem 5.2. Let \( D \in B(S \times A) \) and \( x \in S \) be arbitrary but fixed, and suppose that

\[
M_{\pi}(x, y, D, \varepsilon) < 1, \quad y \in S.
\]

In this case, there exists a positive constant \( \varepsilon \) such that

\[
M_{\pi}(x, y, D, \varepsilon) < 1, \quad y \in S.
\]

The proof relies on the following two lemmas.

Lemma 5.1. Let \( D \in B(S \times A) \) and \( \tilde{M} S \to (0, \infty) \) be given and suppose that

\[
\tilde{M}(x) \geq e^{-e(D)\sum_{y} P_{y}(a)M_{y}(y), \quad (x, a) \in S \times A.}
\]

In this case, \( M_{\pi}(x, z, D) \leq 1, \quad x \in S \).

Proof. Define \( M_{0} = \tilde{M}(X_{0}) \) and \( M_{n} = \sum_{i=0}^{n-1} e^{-e(D)M_{i}} \tilde{M}(X_{i+1}) \) for \( n \geq 1 \) and, as before, let \( P_{\pi}^{n} \) be the \( n \)-field generated by the history vector \( H_{n} = (X_{0}, A_{0}, \ldots, X_{n-1}, A_{n-1}, h_{n}) \). Notice now that for arbitrary \( x \in S \) and \( \pi \in P \)

\[
E_{\pi}^{y}[M_{n+1} | F_{n}] = E_{\pi}^{y}[\sum_{i=0}^{n+1} e^{-e(D)M_{i}} \tilde{M}(X_{i+1}) | F_{n}] 
\]

\[
= e^{-e(D)M_{n+1}} E_{\pi}^{y}[e^{-e(D)M_{n+1}} \tilde{M}(X_{n+1}) | F_{n}] 
\]

\[
\leq e^{-e(D)M_{n+1}} E_{\pi}^{y}[\sum_{i=0}^{n+1} e^{-e(D)M_{i}} \tilde{M}(X_{i+1}) | F_{n}] 
\]

where the shifted policy \( \pi^{*} \) is defined by \( \pi^{*} \{ h_{k} \} = \pi^{*}(x, A_{0}, \ldots, X_{k-1}, A_{k-1}, h_{k}) \), and the Markov property was used to obtain the second equality. Therefore, (5.5) yields that

\[
E_{\pi}^{y}[M_{n+1} | F_{n}] \leq e^{-e(D)M_{n}} M_{n+1}
\]

so that \( M_{n} \) is a submartingale with respect to each probability measure \( P_{\pi}^{n} \) and the filtration \( \{ F_{n} \} \). By optional stopping, for each positive integer \( n \),

\[
\tilde{M}(x) = E_{\pi}^{y}[M_{n}]
\]

\[
\geq E_{\pi}^{y}[M_{n+1}]
\]

\[
= e^{-e(D)M_{n+1}} E_{\pi}^{y}[e^{-e(D)M_{n+1}} \tilde{M}(X_{n+1}) | F_{n}]
\]

\[
= e^{-e(D)M_{n+1}} E_{\pi}^{y}[\sum_{i=0}^{n+1} e^{-e(D)M_{i}} \tilde{M}(X_{i+1}) | F_{n}]
\]

\[
= e^{-e(D)M_{n+1}} \sum_{i=0}^{n+1} e^{-e(D)M_{i}} \tilde{M}(X_{i+1}) | F_{n} \leq n
\]

\[
= \sum_{i=0}^{n} e^{-e(D)M_{i}} \tilde{M}(x) | F_{n} \leq k
\]
where it was used that $X_\epsilon = x$ on the event $[\mathcal{T}_\epsilon = k]$. Observe now that
\[ P_k^\epsilon \mathbb{P}[\mathcal{T}_\epsilon < \infty] = 1 \] (by Theorem 4.1), so that letting $n$ increase to $\infty$ in the last inequality, it follows that
\[ \hat{M}(x) \geq \sum_{k=1}^{\infty} \mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \hat{M}(x)[\mathcal{T}_\epsilon = k] \right] = M(x) \mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \right]. \]
and then, using the condition $\hat{M}(x) > 0$, this yields that $\mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \right] \leq 1$, and the conclusion follows, since $e$ and $\mathbb{P}$ were arbitrary.

Lemma 5.2.1: Let $D \in \mathcal{B}(S \times A)$ be a borel function and let $x \in S$ be such that,
\[ M_x(x, x, D) < 1. \]

In this case
(i) There exists $d > 0$ such that $M_x(x, x, D + d1_k) < 1$.

Moreover,
(ii) $M_x(y, x, D + d1_k) < 1$ for every $y \notin S$.

Proof. Let $x \in S$ be such that $M_x(x, x, D) < 1$.

(i) The starting point is the optimality equation in Theorem 5.1(ii):
\[ M_x(y, x, D) = \sup_{a} \left\{ e^{D(y,a)} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w, x, D) \right) \right\}, y \in S. \tag{5.6} \]

Next, define $\hat{M} = \hat{M} \in \mathcal{B}(\mathcal{E})$ and $\hat{D} \in \mathcal{B}(S \times A)$ by
\[ \hat{M}(y) = M_x(y, x, D), \quad y \neq x, \quad \text{and} \quad \hat{M}(x) = 1, \tag{5.7} \]
\[ \hat{D}(x) = d + 2d1_k. \tag{5.8} \]

where
\[ 2d = - \log(M_x(x, x, D)) > 0. \tag{5.9} \]

Combining these definitions with (5.6), it is clear that
\[ \hat{M}(y) = \sup_{a} \left\{ e^{\hat{D}(y,a)} \sum_{u \notin S} p_{w|u} M_x(w, x, D) \right\}, \text{for } y \neq x, \tag{5.10} \]

whereas for $y = x$, (5.6)-(5.9) yield
\[ \hat{M}(x) = \sum_{k=1}^{\infty} \mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \hat{M}(x)[\mathcal{T}_\epsilon = k] \right] = M(x) \mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \right]. \]

Next, observe that
\[ \hat{D}(X_k, A_k) \geq D(X_k, A_k) + d1_k(X_k) \text{ and } \hat{D}(x, A_0) \geq d + [D(x, A_0) + d1_k(x)] \] (see (5.8)), so that, setting $y = x$ in (5.11),
\[ 1 \geq M_x(x, x, \hat{D}) = \sup_{a} \left\{ e^{\hat{D}(y,a)} \sum_{u \notin S} p_{w|u} M_x(w, x, D) \right\}, \tag{5.11} \]

\[ \geq \sup_{a} \mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \right] \]

\[ \geq e^{\delta \sup_{a} \mathbb{E}^\epsilon \left[ e^{\sum_{i=0}^{N_k-1} \delta(N_k, a_k)} \right] \]

\[ = e^{\delta M_x(x, x, D + d1_k)} \]

and then $1 > e^{-d} \geq M_x(x, x, D + d1_k)$.

(ii) First, notice that the inequality in (5.12) yields $M_x(y, y, \hat{D})$. 

Risk-sensitive average cost optimality equation in communicating MDP's 
\[ \hat{M}(x) = 1 \]
\[ = \sup_{a} \left\{ e^{\hat{D}(x,a)} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w, x, D) \right) \right\} \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a) + d1_k} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a)} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a)} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a)} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a)} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a)} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
\[ \geq \sup_{a} \left\{ e^{\hat{D}(x,a)} \right\} \left( P_{x|a} + \sum_{u \notin S} p_{w|u} M_x(w) \right) \]
by Jensen’s inequality. Next, observe that

\[
\int LdQ
\]

\[
= \frac{1}{P_y(T_x \leq T_y)} \int_{T_x \leq T_y} L dP_y
\]

\[
= \frac{1}{P_y(T_x \leq T_y)} \int_{T_x \leq T_y} \sum_{n=0}^{\infty} (|D| + 1)^n \left( \frac{\Delta}{\Delta + 1} \right)^n dQ
\]

\[
\geq \frac{1}{P_y(T_x \leq T_y)} \int_{T_x \leq T_y} \sum_{n=0}^{\infty} (|D| + 1)^n \left( -\frac{(|D| + 1)K}{\Delta} \right) dQ
\]

where \( K \) is as in Theorem 4.1, and combining this inequality with (5.14) and (5.15) it follows that

\[
M_\varepsilon(y, D, \bar{D}) - M_\varepsilon(y, D + d_k) \geq P_y(T_x < T_y) \exp \left( \frac{(|D| + 1)K}{\Delta} \right)
\]

To conclude observe that the mapping \( b \mapsto \theta \exp\left( -\left( \frac{(|D| + 1)K}{\Delta} \right) b \right) \) is increasing in \( b \in (0, \infty) \), so that, the last inequality yields

\[
M_\varepsilon(y, D, \bar{D}) - M_\varepsilon(y, D + d_k) \geq \Delta \exp \left( -\left( \frac{(|D| + 1)K}{\Delta} \right) \right)
\]

where \( \Delta > 0 \) is as in Theorem 4.2. Then,

\[
M_\varepsilon(y, D, \bar{D}) \leq M_\varepsilon(y, D) - \Delta \exp \left( -\left( \frac{(|D| + 1)K}{\Delta} \right) \right)
\]

so that, by (5.1), (5.2) and (5.11),

\[
M_\varepsilon(y, D + d_k) \leq M_\varepsilon(y, D) - \Delta \exp \left( -\left( \frac{(|D| + 1)K}{\Delta} \right) \right) \leq 1 - \Delta \exp \left( -\left( \frac{(|D| + 1)K}{\Delta} \right) \right) < 1.0
\]
After these preliminaries, Theorem 5.2 can now be proved.

**Proof of Theorem 5.2.** Let $x \in S$ be such that $M_x(z, x; D) < 1$, and write $S = \{x_1, x_2, \ldots, x_k\}$, where $x = x_1$ and $N$ is the number of elements of $S$. Given a positive integer $k \leq N$, set $S_k = \{x_1, \ldots, x_k\}$ and let $I_{S_k}$ denote the indicator function of $S_k$, i.e.,

$$I_{S_k}(y) = 1, \quad y \in S_k, \quad I_{S_k}(y) = 0, \quad y \in S \setminus S_k. \tag{5.16}$$

For $1 \leq k \leq N$, consider the following claim:

There exists a positive constant $c_k$ such that $M_x(y, y; D + c_k I_{S_k}) < 1$, $y \in S$.

(a) Applying Lemma 5.2 with $z = x_k$, (5.16) hold true for $k = 1$.

(b) Suppose that (5.16) is valid for $k = r < N$. In this case, from an application of Lemma 5.2 with $x_{r+1}$ instead of $x$ and with $D + c_r I_{S_r}$ replacing $D$, it follows that there exists $d > 0$ such that for every $y \in S$,

$$M_x(y, y; D + c_r I_{S_r} + dI_{S_{r+1}}) < 1, \quad y \in S. \tag{5.6}$$

Then, setting $c_{r+1} = \min\{c_r, d\} > 0,

$$M_x(y, y; D + c_{r+1} I_{S_{r+1}}) \leq M_x(y, y; D + c_r I_{S_r} + dI_{S_{r+1}}) < 1, \quad y \in S,$$

so that (5.16) holds true for $k = r + 1$.

From (a) and (b), it follows that (5.16) is valid for $k = N$, that is, for $c = c_N > 0,

$$M_x(y, y; D + c) = M_x(y, y; D + c_N I_{S_N}) < 1 \text{ for all } y \in S. \tag{5.17}$$

6. **AUXILIARY EXPECTED–TOTAL COST PROBLEMS: II**

In this section, by examining (3.1), (5.1-5.2), and by an appropriate choice of the one-stage cost function $B(\cdot, \cdot)$ in the auxiliary expected–total cost MDP candidate solutions $g_n, h_n$ needed to establish Theorem 3.1 are constructed.

**Theorem 6.1.** Let $z \in S$ and $\lambda \neq 0$ be arbitrary but fixed.

(i) There exists $g^*_2 \in \mathcal{R}$ such that $M_x(z, z; \lambda(C - g^*_2)) = 1$.

Similarly,

(ii) There exists $g^*_2 \in \mathcal{R}$ such that $M_x(z, z; \lambda(C - g^*_2)) = 1$.

The proof of this result is presented in two parts below, as Lemmas 6.1 and 6.2. First, it is convenient to introduce some useful notation.
Lemma 6.2. Let \( z \in S \) and \( \lambda \neq 0 \) be fixed.

(i) For each \( g \in G_-(\lambda, z) \), there exists \( f \in \mathcal{F} \) such that

\[
M_-(y, z, \lambda C + g) = M_j(y, z, \lambda C + g), \quad y \in S.
\]

Let \( g_{z, \lambda} = \sup\{g \in G_-(\lambda, z)\} \). In this case,

(ii) \( g_{z, \lambda} \in G_-(\lambda, z) \), and

(iii) \( M_-(z, z, \lambda C + g_{z, \lambda}) = 1 \).

Proof. (i) Notice that \( M_-(y, z, \lambda C + g) \leq 1 \), since \( g \in G_-(\lambda, z) \), so that Theorem 5.1(iii) yields that

\[
\infty > M_-(y, z, \lambda C + g) = \inf \left\{ \mathbb{E}^\pi \left[ \mathbb{E}^\mathbb{F} \left[ \sum_{n=0}^{\infty} p_{y,z}(a) \mathbf{1}_{\{g_{z,\lambda} = g + \alpha\}} \mathbb{E}^\mathbb{F} \left[ \sum_{n=0}^{\infty} p_{y,z}(a) M_-(w, z, \lambda C + g) \right] \right] \right] : y \in S \right\}
\]

and then, by Assumption 2.1, there exists a policy \( f \in \mathcal{F} \) such that for every \( y \in S \),

\[
M_-(y, z, \lambda C + g) = \mathbb{E}^\pi \left[ \sum_{n=0}^{\infty} p_{y,z}(a) \mathbf{1}_{\{\text{exists } n \text{ s.t. } \mathbb{E}^\mathbb{F} \left[ \sum_{n=0}^{\infty} p_{y,z}(a) M_-(w, z, \lambda C + g) \right] = 1}\} \right] = 1.
\]

(ii) From this equality, a simple induction argument using the Markov property yields that for every positive integer \( n \) and \( y \in S \),

\[
M_-(y, z, \lambda C + g) = \sum_{k=1}^{n} \mathbb{E}^f \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} \mathbb{E}^\mathbb{F} \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} M_-(x, x, \lambda C + g) \right] \right] + \mathbb{E}^f \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} M_-(x, x, \lambda C + g) \right] \mathbb{P}^f \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} M_-(x, x, \lambda C + g) \right] = \sum_{k=1}^{n} \mathbb{E}^f \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} \right],
\]

and since \( \mathbb{E}^f \left[ T_k < \infty \right] = 1 \), this last inequality implies, for each \( y \in S \), that

\[
M_-(y, z, \lambda C + g) = \lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{E}^f \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} \right] = \mathbb{E}^f \left[ \sum_{k=1}^{n} \mathbf{1}_{\{T_k = k\}} \right] = 1.
\]
(iii) By parts (i) and (ii), there exists a policy \( f \in \mathbb{F} \) such that
\[
1 \geq M_f (z, \rho, \lambda) = M_f (z, \rho, \lambda) + g_{\lambda}(z) = 1 \geq M_f (z, \rho, \lambda) + g_{\lambda}(z).
\]
It will be shown, by contradiction, that \( M_f (z, \rho, \lambda) + g_{\lambda}(z) = 1 \). Thus, suppose that
\[
M_f (z, \rho, \lambda) + g_{\lambda}(z) = M_f (z, \rho, \lambda) < 1, \tag{6.5}
\]
and define a modified MDP \( M = (S, A, \{p_x \}_{x \in S}) \). By setting \( p_x (a) = p_x (f(x)) \) and \( C(x, a) = C(x, f(x)) \). In this case, using (6.5), Lemma 5.2 applied to model \( M \) yields a positive constant \( c \) such that
\[
M_f (z, \rho, \lambda) + g_{\lambda}(z) < c < 1,
\]
and thus \( M_f (z, \rho, \lambda) + g_{\lambda}(z) < 1 \). This contradicts the definition of \( g_{\lambda}(z) \), as the supremum of \( C(\lambda, z) \). Therefore \( M_f (z, \rho, \lambda) + g_{\lambda}(z) = 1 \) and the proof is complete.

The two previous lemmas yield Theorem 6.1 immediately.

**Proof of Theorem 6.1.** Let \( \lambda \neq 0 \) be fixed. Setting \( g_{\lambda}(z) = -g_{\lambda}(z) \), part (i) follows from Lemma 6.1, whereas defining \( g_{\lambda}(z) = -g_{\lambda}(z) \), Lemma 6.2 yields part (ii).

7. PROOF OF THEOREM 3.1

After the basic technical results presented in the previous sections, Theorem 3.1 can be established as follows.

**Proof of Theorem 3.1.** Let \( \lambda \neq 0 \) and \( z \in S \) be fixed, and define \( (h_\lambda, h_\lambda(z)) \in \mathbb{R} \times \mathcal{B}(S) \) as follows:

For \( \lambda < 0 \):
\[
h_\lambda(z) = g_{\lambda} \left[ M_f (z, \rho, \lambda) \right], \quad y \in S, \tag{7.1}
\]
whereas

For \( \lambda > 0 \):
\[
h_\lambda(z) = g_{\lambda} \left[ M_f (z, \rho, \lambda) \right], \quad y \in S. \tag{7.2}
\]

By Theorem 6.1, (7.1) and (7.2) yield that \( h_\lambda(z) = 0 \), and then, the optimality equations in Theorem 5.1 imply that
\[
\rho^{\lambda,(\cdot)} = \sup_y \left\{ e^{C(\lambda, y, \lambda)} \sum_x p_x (f(x)) e^{h_\lambda(z)} \right\}, \quad y \in S, \lambda < 0,
\]
and
\[
\rho^{\lambda,(\cdot)} = \inf_y \left\{ e^{C(\lambda, y, \lambda)} \sum_x p_x (f(x)) e^{h_\lambda(z)} \right\}, \quad y \in S, \lambda > 0.
\]

Risk-sensitive average cost optimality equation in communicating MDP's

equalities that can be condensed into a single one: For \( \lambda \neq 0 \),
\[
\rho^{(\lambda)}(z) = \inf_y \left\{ e^{C(\lambda, y, \lambda)} \sum_x p_x (f(x)) e^{h_\lambda(z)} \right\}, \quad y \in S,
\]
which is equivalent to the \( \lambda \)-ACOE (3.1).

(i) and (iii) These parts can be obtained as in Theorem 3.1 in Cavazos-Cadena and Fernández-Gaucherand (1989a), or from the verification theorem in Hernández-Hernández and Marcus (1996) for the case \( \lambda > 0 \).

(ii) Let \( (h_\lambda, h_\lambda(z)) \in \mathbb{R} \times \mathcal{B}(S) \) be such that
\[
\rho^{(\lambda)}(z) = \inf_y \left\{ e^{C(\lambda, y, \lambda)} \sum_x p_x (f(x)) e^{h_\lambda(z)} \right\}, \quad y \in S, \lambda > 0,
\]
where \( h_\lambda(z) = h_\lambda(z) \), for some fixed state \( z \in S \). It will be shown that \( h_\lambda(z) = h_\lambda(z) \) and that \( h_\lambda(z) = h_\lambda(z) \). First, notice that from Theorem 3.1 in Cavazos-Cadena and Fernández-Gaucherand (1989a) it follows that \( h_\lambda(z) = h_\lambda(z) \) so that \( g_{\lambda}(z) \) is the optimal average cost at every state, which is uniquely determined, so that \( g_{\lambda}(z) = g_{\lambda}(z) \). Next, using Assumption 2.1, select \( f \in \mathbb{F} \) such that \( f(x) \) is a minimizer of the term in brackets in (7.3), so that
\[
\rho^{(\lambda)}(z) = e^{C(\lambda, y, \lambda)} \sum_x p_x (f(x)) e^{h_\lambda(z)}, \quad y \in S, \lambda > 0. \tag{7.3}
\]

In this case, Lemma 4.1 with \( h_\lambda(z) \) and \( \lambda C(\lambda, f(x)) - g_\lambda(z) \) instead of \( h(\cdot) \) and \( C(\cdot) \), respectively, yields that for every \( z \in S \) and \( i = 1, 2 \),
\[
\rho^{(\lambda)}(z) = \lim_{n \to \infty} \mathbb{E}^f \left[ \sum_{t=0}^{n-1} \lambda C(\lambda, X_t, \lambda) - g_\lambda(z) \right| T_i \leq n \] \tag{7.4}

(notice that \( h(X_0, \lambda) = h_\lambda(z) = 0 \) on the event \( T_i > n \)), and
\[
\lim_{n \to \infty} \mathbb{E}^f \left[ \sum_{t=0}^{n-1} \lambda C(\lambda, X_t, \lambda) - g_\lambda(z) \right| T_i > n \] \tag{7.5}

To conclude, consider the following two cases.

Case 1: \( \lambda > 0 \). Since \( g_\lambda(z) = g_{\lambda}(z) \), in this situation (7.3) implies that \( \rho^{(\lambda)}(z) \leq \mathbb{E}^f \left[ e^{C(\lambda, X_0, \lambda) - g_\lambda(z)} \lambda X_1 \right] \), and following the same arguments as in the proof of Lemma 4.1(i),
\[
\rho^{(\lambda)}(z) \leq \mathbb{E}^f \left[ \sum_{t=0}^{\infty} \lambda C(\lambda, X_t, \lambda) - g_\lambda(z) \right| X_0 = x, \lambda > 0. \tag{7.6}
\]
and then, using that \( h_1(x_T) = h_1(x) = 0 \),

\[
\begin{align*}
\mathcal{A}^{(1)}_T &\leq E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T \leq n) \right] \\
&+ E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T > n) \right] \\
&\leq E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T > n) \right] \\
&+ \epsilon^n(1) E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T > n) \right] \\
&= \epsilon^n(1) E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T > n) \right] \tag{7.7}
\end{align*}
\]

and (7.4) and (7.5) imply that

\[
\begin{align*}
\mathcal{A}^{(2)}(x) &\leq \lim_{n \to \infty} E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T \leq n) \right] \\
&+ \epsilon^n(1) \lim_{n \to \infty} E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T > n) \right] \\
&= \epsilon^n(1) = 0, \quad \mathcal{A}^{(2)}(x) \tag{7.8}
\end{align*}
\]

so that \( h_2(x) = h_2(x) \); the reverse inequality can be obtained in a similar way.

and then \( h_1(\cdot) = h_2(\cdot) \).

**Case 2: \( \lambda < 0 \).** Using again that \( \mu = g_\lambda \); (7.3) yields \( \mathcal{A}^{(1)}_T \geq E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n}) \right] \) for every \( x \in S \), and then (7.0)–(7.8) occur if the inequalities are reversed and in (7.7) \( \epsilon^n(1) \) is replaced by \( \epsilon^n(1) \). Therefore \( \mathcal{A}^{(1)}_T \geq E^P \left[ \sum_{n=0}^{T} \lambda(C(X_{t+n})-\mu)A(X_{t+n})I(T_T > n) \right] \).

so that \( h_1(\cdot) \geq h_2(\cdot) \) and, similarly, \( h_2(\cdot) \leq h_1(\cdot) \), so that \( h_1(\cdot) = h_2(\cdot) \).

\( \square \)

8. CONCLUSIONS

This paper considered Markov decision processes endowed with the risk-sensitive average cost optimality criterion in (2.4)–(2.6). The main result of the paper, namely, Theorem 3.1, shows that under standard continuity–compactness conditions (as in Assumption 2.1), the communication condition in Assumption 2.3 guarantees the existence of a solution to the \( \lambda_\lambda \)-ACOE stated in (3.1) for arbitrary values of \( \lambda \) \( > 0 \), when the state space is finite. Furthermore, it was shown via Example 3.1, that the conclusions in Theorem 3.1 cannot be extended to the case of countably infinite state space models. Hence, the results presented in the paper significantly extend those in Howard and Matherson (1972), and also the recent work of the authors presented in Cavazos-Cadena and Fernández-Gaucherand (1998a–b), see Remark 3.1.

**REFERENCES**


REFERENCES


APPENDIX: A: PROOF OF THEOREM 4.1

Theorem 4.1 stated in Section 4 is a consequence of the following.

Theorem A. Let the state space be finite suppose that Assumptions 2.1 and 2.3 hold true. In this case, the uniqueness condition is valid. More explicitly, there exist $\epsilon \in (0, \infty)$ such that $E[S] \leq K$, $\alpha \in S$, $f \in F$.

Notice that the conclusion of Theorem A refers to the class $F$ of stationary policies, whereas Theorem 4.1 involves the family of all policies. However, Theorem 4.1 can be derived from Theorem A by taking $F$ in the condition $F$ to be the class of all policies $\pi$ that satisfy $\pi(x) = 0$ for all $x \in S$.
MODELING UNCERTAINTY

Using the S is finite, it is clear that \( a() \geq 0 \) and \( \sum_{x} p(y) = 1 \), since each \( \mu_{x} \) has the properties: monotone, for each \( x \in S \).

\[
\begin{align*}
\mu(x) &= \lim_{n \to \infty} \mu_n(x) = \lim_{n \to \infty} \sum_{y} p(y) f_n(x, y) \\
&= \sum_{y} \lim_{n \to \infty} p(y) f_n(x, y) \\
&= \sum_{y} p(y) f(x, y)
\end{align*}
\]

where the last equality used \( (A, 2) \) and Assumption 2.1. Therefore, \( \mu(x) \) has all the properties to be the unique invariant distribution of the transition maps determined by \( f \), so that \( \mu = \mu_f \).

(iii) As already noted, \( \mu_{x}(x) > 0 \), by Assumption 2.3, so that using the equality \( \mu_{x} = f_{x}(x, y) \) (Loeb, 1980), the assertion follows from part (i).

Lemma A.2. Let \( f \in \mathcal{F} \) and \( x, y \in S \) with \( x \neq y \) be arbitrary.

(i) There exists a positive integer \( k \) such that \( P_{f}^{k}(x, y) > 0, P_{f}^{k}(x, y) > 0 \).

(ii) For each \( x, y \in S \), \( E_{f}^{x}(x, y) \) is finite.

Proof. (i) By Assumption 2.3, there exists a positive integer \( m \) and states \( x, x_1, \ldots, x_n = y \) such that

\[
\sum_{s=1}^{n} \mu_{s}(x) > 0.
\]

Let \( j = \min \{ s \mid x_s = x \} \), so that \( j < m \). Let \( m = m - j > 0 \) and notice that \( x_j, x_{j+1}, \ldots, x_n = y \). Next, set \( k = m \) and notice that \( \mu_{x} \neq \mu_{x} \) for \( x \neq y \), so that (see (A.3))

\[
\begin{align*}
P_{f}^{k}(x, y) &> \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0
\end{align*}
\]

(ii) Let \( k \) as in part (i), and notice that

\[
\begin{align*}
P_{f}^{k}(x, y) &> \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0
\end{align*}
\]

Therefore, denoting the distribution generated by \( f_{x}, \ldots, f_{x} \) by \( f_{k} \), the Markov property yields:

\[
\begin{align*}
E_{f}^{x}(x, y) &> \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0
\end{align*}
\]

REFERENCES

Combining (A.2), (A.5), it follows that

\[
\begin{align*}
E_{f}^{x}(T_{f}^{k}) &> \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0
\end{align*}
\]

Then, Lebesgue, 1980, it follows that \( E_{f}^{x}(T_{f}^{k}) < \infty \).

Lemma A.3. Let \( x \in S \) be fixed and let \( D \in \mathcal{B}(S) \) be the indicator function of \( S \setminus \{ x \} \), i.e.,

\[
D(x) = 1, \quad D(x) = 1
\]

and set

\[
\begin{align*}
E[x] &= \sup_{x \in S} \left( E_{f}^{x}(\sum_{t=1}^{\infty} D(x)) \right) \\
\end{align*}
\]

Then

\[
\begin{align*}
E[x] &= \sup_{x \in S} \left( E_{f}^{x}(\sum_{t=1}^{\infty} D(x)) \right) \\
\end{align*}
\]

(iii) Define \( h \in \mathcal{F} \) by

\[
\begin{align*}
h(y) &= \sup_{x \in S} \left( E_{f}^{x}(\sum_{t=1}^{\infty} D(x)) \right) \\
\end{align*}
\]

Then, \( h(x) = 1 \) and the following non-stochastic optimality equation is satisfied:

\[
\begin{align*}
g(x) \leq \sup_{y \in S} \left( E_{f}^{x}(\sum_{t=1}^{\infty} D(y)) \right) \\
\end{align*}
\]

Proof. (i) Since \( D(x) = 0 \) and \( D(x) = 1 \), it is clear that

\[
\begin{align*}
E_{f}^{x}(\sum_{t=1}^{\infty} D(x)) = E_{f}^{x}(\sum_{t=1}^{\infty} 1) = E_{f}^{x}(1) = 1
\end{align*}
\]

Therefore, denoting the distribution generated by \( f_{x}, \ldots, f_{x} \) by \( f_{k} \), the Markov property yields:

\[
\begin{align*}
E_{f}^{x}(x, y) &> \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0
\end{align*}
\]

\[
\begin{align*}
E_{f}^{x}(x, y) &> \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0 \\
&= \sum_{s=1}^{n} \mu_{s}(x) > 0
\end{align*}
\]

(4.4)

Then

\[
\begin{align*}
E[x] &= \sup_{x \in S} \left( E_{f}^{x}(\sum_{t=1}^{\infty} D(x)) \right) \\
\end{align*}
\]

(4.5)
MODELING UNCERTAINTY

(4.1) Notice that the equality \( h(y) = 0 \) follows from (4.5). Also, using the Markov property, it is not difficult to see that for every \( y \in S \)

\[ h(y) = D(y) - g + \sum_{x \in X} p_x f_x(y) h(x) = D(y) - g + \sum_{x \in X} p_x f_x(y) h(x) \]

which is equivalent to

\[ h(y) + g = D(y) + \sum_{x \in X} p_x f_x(y) h(x) \quad y \in S \quad (4.8) \]

To establish (4.6), pick an arbitrary pair \((u, a) \in S \times A\) and define the discrepancy function \( \Phi \in \mathcal{B}(S) \) and the policy \( f \in \mathcal{F} \) by

\[ \Phi(y) = \begin{cases} (y - u - D(u) - \sum_{x \in X} p_x h(x)) & 
\text{if } y = u, \\
0 & \text{otherwise.} \end{cases} \quad (4.9) \]

and

\[ f(a) = u \quad f(y) = f_x(y) \quad y \neq u. \]

Combining these definitions with (4.8) it follows that for every \( y \in S \)

\[ h(y) + g = D(y) + \Phi(y) + \sum_{x \in X} p_x f_x(y) h(x) \]

and then

\[ y = \lim_{n \to +\infty} \frac{1}{n+1} \sum_{s=1}^{n} D(X_s) + \Phi(X_s) \]

\[ = \sum_{x \in X} \mu_x(D(x) + \Phi(x)) = \sum_{x \in X} \mu_x(D(x) + \sum_{x \in X} p_x h(x)) \]

Combining this last equality with (4.7) and using the fact that \( \mu_x(x) > 0 \), it follows that

\[ \Phi(y) = 0. \]

Therefore (see (4.9), (4.10), (4.11), (4.12)) and then, since the pair \((u,a) \in S \times A\) was arbitrary,

\[ h(y) = \sup_{x \in X} \left[ D(x) + \sum_{x \in X} p_x h(x) \right]. \]

and (4.8) follows combining the inequality with (4.8).

Proof of Theorem A. The notation is as in the statement of Lemma A.3. From the optimality equation (4.6), and using that \( h(y) = 0 \), it follows that for every \( f \in \mathcal{F}\) and \( y \in S \),

\[ h(y) = D(y) - g + \sum_{x \in X} p_x f_x(y) h(x) = D(y) - g + \sum_{x \in X} p_x f_x(y) h(x) \]

and an induction argument yields that, for every \( n \in \mathbb{N} \),

\[ h(y) \geq E^f_n \left[ \sum_{x \in X} (D(X_n) - g) I[T_n > t] \right] + E^f_n [h(X_{n+1}) I[T_n > t]] \quad (4.10) \]

REFERENCES

Observe now that \( D(X_0) - g \geq 0 \) whereas for positive \( t, T_t \geq t \Rightarrow X_t \neq x \Rightarrow D(X_t) - g = 1 - g \), so that (A.10) implies that

\[ 2D(y) + g \geq E^f_{n+1} \left[ \sum_{t=1}^{\infty} (1 - g) I[T_t > t] \right] = (1 - g) \sum_{t=1}^{\infty} P[T_t > t] \]

and letting \( n \) increase to \( \infty \), it follows that

\[ 2D(y) + g \geq (1 - g) \sum_{t=1}^{\infty} P[T_t > t] = \left( 1 - g \right) E^f T_t = \left( 1 - g \right) E^f T_t - 1 \]

since \( y \) is strictly less than \( y \) by Lemma A.2 (i). \( E^f T_t = 1 \Rightarrow 1 + 2D(y) + g = 1 - g E^f T_t - 1 \), so that

\[ \sup_{a \in A} E^f T_t = K < \infty \]

To conclude observe that in this argument \( T \) is \( \mathcal{F}_{\infty} \) stationary, so that the setting \( K = \max \left\{ E^f T_t \mid t \in S \right\} < \infty \),

\[ \sup_{a \in A} E^f T_t = K < \infty \]

APPENDIX B: PROOF OF THEOREM 4.2

The proof of Theorem 4.2 relies on the following lemma.

Lemma B.1. Suppose that Assumptions 2.1 and 2.3 hold true. For given \( x, y \in S \), the number of times of the state process \( \{x\} \) before returning to the initial state \( y \) in a positive time, is

\[ N(x, y) = \sum_{t=1}^{\infty} I_{x_t=y} \]

is finite, and

\[ E^f N(x, y) \leq K P[T_1 \leq T_2], \]

where \( K \) is as in Theorem 4.1. [Recall that \( I_{x_t=y} \) is the indicator function of \( \{x_t=y\} \).]

Proof. For each positive \( n \), notice that given \( x_0 = x \)

\[ E^f N(x, y) = E^f \sum_{t=1}^{n} I_{x_t=y} \]

\[ = E^f \left[ x_0 = y, y < x, x_t = y \right] \sum_{t=1}^{n} I_{x_t=y} \]

\[ = E^f \left[ x_0 = y, y < x, x_t = y \right] \sum_{t=1}^{n} I_{x_t=y} \]

\[ = E^f \left[ x_0 = y, y < x, x_t = y \right] \sum_{t=1}^{n} I_{x_t=y} \]

in that, setting \( P_1 \) be the drift generated by \( H_1 \), the Markov property yields

\[ E^f [X_{n+1} = x, y < x, x_t = y] = E^f [x_0 = y, x_t = y] \]

\[ = E^f \left[ x_0 = y, y < x, x_t = y \right] \sum_{t=1}^{n} I_{x_t=y} \]

\[ = E^f \left[ x_0 = y, y < x, x_t = y \right] \sum_{t=1}^{n} I_{x_t=y} \]
where the shifted policy \( \pi' \) is determined by \( \pi'(x) = \pi(x + k), k \in \mathbb{Z} \), and

\[
E_p' \Pi(x, y, s, k, X_k = x) = \mathbb{E}_p' \left[ \sum_{\tau=0}^{\infty} r(X_{\tau}) \right]
\]

where \( K \) is as in Theorem 4.1. Hence, taking expectation with respect to \( p' \), it follows that

\[
E_p' \Pi(x, y, s, k, X_k = x) = \mathbb{E}_p' \left[ \sum_{\tau=0}^{\infty} r(X_{\tau}) \right] \leq K \mathbb{E}_p' \left[ e^{\gamma \tau} \right]
\]

and then

\[
E_p' \Pi(x, y, s, k, X_k = x) = \mathbb{E}_p' \left[ \sum_{\tau=0}^{\infty} r(X_{\tau}) \right] \leq 2K \mathbb{E}_p' \left[ e^{\gamma \tau} \right]
\]

which, using that \( \mathbb{E}_p' \left[ e^{\gamma \tau} \right] \leq 1 \) on the event \( \{T > T_p\} \), yields

\[
E_p' \Pi(x, y, s, k, X_k = x) \leq 2K \mathbb{E}_p' \left[ e^{\gamma \tau} \right]
\]

To conclude, observe that given \( X_0 = y, N(x, y) = 1 \) on the event \( \{T > T_p\} \), so that

\[
E_p' \Pi(x, y, s, k, X_k = x) \leq 2K \mathbb{E}_p' \left[ e^{\gamma \tau} \right] = 2K \mathbb{E}_p' \left[ e^{\gamma \tau} \right]
\]

Proof of Theorem 4.2. (i) Let \( x, y \in S \) be given with \( x \neq y \). Since the simultaneous DCR condition holds, there exists \( g \in \mathcal{R} \) and \( \epsilon \in \mathbb{B}(S) \) such that \( \mathbb{E}[g, \epsilon] \geq a \). (Jahreise, 1972; (Jahreise, 1976))

\[
g + h(x) = \inf_{u} \left[ E_p' [u] + \sum_{x \in S} \mu(x, e) h(x) \right], \quad w \in S.
\]

where without loss of generality, it is assumed that

\[
h(\pi) \equiv 0 \quad (B.1)
\]

If for each \( w \in S, \pi \in \Pi \) is such that \( f(w) \) minimizes the term in brackets in (B.1), it follows that

\[
g = \lim_{n \to \infty} \frac{1}{n+1} \sum_{\tau=0}^{n} r(x_{\tau}) = \sum_{k=1}^{n} \frac{1}{n+1} r(x_{\tau}) = \mathbb{E}_p' \left[ r(X_{\tau}) \right]
\]

On the other hand, (B.1) implies that for each \( x \in S \) and \( \pi \in \Pi ,

\[
h(\pi) \leq E_p' [g(X_t)] - g + h(x)] - \mathbb{E}_p' [r(X_{\tau})] \geq h(x) - |\mathbb{E}_p' [r(X_{\tau})]| \geq 0
\]

where the equality used that \( h(\pi) = 0 \) as well as \( E_p' [r(X_{\tau})] = 0 \). Next, an induction argument using the Markov property, yields that for each positive integer \( n, w \in S \), and \( \pi \in \Pi ,

\[
E_p' [g(x)] \leq E_p' \left[ \sum_{\tau=0}^{\infty} r(X_{\tau}) + h(x) \left( T_{\pi} > T_p \right) + h(x) \left( T_{\pi} \leq T_p \right) \right]
\]

Observe now that \( \sum_{\tau=0}^{\infty} r(X_{\tau}) + h(x) \left( T_{\pi} \leq T_p \right) \leq T_p + |\mathbb{A}| \), and since \( E_p' [g(x)] < \infty \)

(by Theorem 4.1), \( P^p \left[ T_{\pi} < \infty \right] = 1 \), so that in \( n \to \infty \), the following convergences hold \( P^p \)-almost surely:

\[
\sum_{\tau=0}^{\infty} r(X_{\tau}) + h(x) \left( T_{\pi} \leq T_p \right) \to \sum_{\tau=0}^{\infty} r(X_{\tau})
\]

and

\[
h(x) \left( T_{\pi} \leq T_p \right) \to h(x) = 0
\]

Therefore, (B.3) allows to conclude via the dominated convergence theorem, that

\[
h(\pi) \leq \lim_{n \to \infty} \frac{1}{n+1} \sum_{\tau=0}^{n} r(x_{\tau}) + h(x)
\]

Setting \( u = g \) in this inequality follows, since \( h(\pi) = 0 \), that

\[
E_p' [g(x)] \leq \lim_{n \to \infty} \frac{1}{n+1} \sum_{\tau=0}^{n} r(x_{\tau}) + h(x)
\]

and, since \( E_p' [g(x)] \geq 1 \), an application of Lemma 4.1 implies that \( g \leq K \mathbb{E}_p' \left[ e^{\gamma \tau} \right] \), and the conclusion follows obtaining \( x = g/\mathbb{E}_p' \left[ e^{\gamma \tau} \right] > 0 \).

On the other hand, (B.3) implies that for each \( x \in S \) and \( \pi \in \Pi ,

\[
h(\pi) \leq E_p' [g(X_t)] - g + h(x)] - \mathbb{E}_p' [r(X_{\tau})] \geq h(x) - |\mathbb{E}_p' [r(X_{\tau})]| \geq 0
\]

where the equality used that \( h(\pi) = 0 \) as well as \( E_p' [r(X_{\tau})] = 0 \). Next, an induction argument using the Markov property, yields that for each positive integer \( n, w \in S \), and \( \pi \in \Pi ,

\[
E_p' [g(x)] \leq E_p' \left[ \sum_{\tau=0}^{\infty} r(X_{\tau}) + h(x) \left( T_{\pi} > T_p \right) + h(x) \left( T_{\pi} \leq T_p \right) \right]
\]

Observe now that \( \sum_{\tau=0}^{\infty} r(X_{\tau}) + h(x) \left( T_{\pi} \leq T_p \right) \leq T_p + |\mathbb{A}| \), and since \( E_p' [g(x)] < \infty \)

(by Theorem 4.1), \( P^p \left[ T_{\pi} < \infty \right] = 1 \), so that in \( n \to \infty \), the following convergences hold \( P^p \)-almost surely:

\[
\sum_{\tau=0}^{\infty} r(X_{\tau}) + h(x) \left( T_{\pi} \leq T_p \right) \to \sum_{\tau=0}^{\infty} r(X_{\tau})
\]

and

\[
h(x) \left( T_{\pi} \leq T_p \right) \to h(x) = 0
\]

Therefore, (B.3) allows to conclude via the dominated convergence theorem, that

\[
h(\pi) \leq \lim_{n \to \infty} \frac{1}{n+1} \sum_{\tau=0}^{n} r(x_{\tau}) + h(x)
\]

Setting \( u = g \) in this inequality follows, since \( h(\pi) = 0 \), that

\[
E_p' [g(x)] \leq \lim_{n \to \infty} \frac{1}{n+1} \sum_{\tau=0}^{n} r(x_{\tau}) + h(x)
\]

and, since \( E_p' [g(x)] \geq 1 \), an application of Lemma 4.1 implies that \( g \leq K \mathbb{E}_p' \left[ e^{\gamma \tau} \right] \), and the conclusion follows obtaining \( x = g/\mathbb{E}_p' \left[ e^{\gamma \tau} \right] > 0 \).

(a) Let \( x, y \in S \) with \( x \neq y \) given. By (part (i)), \( P^x_1 \leq T_p = T_p \), and using the equality \( \mathbb{E}_p' \left[ g(x) \right] = \mathbb{E}_p' \left[ g(x) \right] \), it follows that \( P^x_1 \left[ g(x) \right] > 0 \) for some positive integer \( k \), and the overall conclusion is obtained observing that \( \{X_k = x \} > 0 \).