TECHNICAL RESEARCH REPORT

Risk Sensitive Control of Markov Processes in Countable State Space

by D. Hernandez-Hernandez, S.I. Marcus

T.R. 96-9

ISR
INSTITUTE FOR SYSTEMS RESEARCH

*Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry*

Risk Sensitive Control of Markov Processes in Countable State Space

Daniel Hernández-Hernández[1] and Steven I. Marcus[2]

Abstract

In this paper we consider infinite horizon risk-sensitive control of Markov processes with discrete time and denumerable state space. This problem is solved proving, under suitable conditions, that there exists a bounded solution to the dynamic programming equation. The dynamic programming equation is transformed into an Isaacs equation for a stochastic game; and the vanishin discount method is used to study its solution. In addtion, we prove that the existence conditions are as well necessary.

Key Words. Risk sensitive control, stochastic dynamic games, Isaacs equation.

[1]Institute for Systems Research, University of Maryland, College Park, Maryland 20742. On leave from Department of Mathematics, CINVESTAV-IPN, MEXICO

[2]Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, Maryland 20742, marcus@src.umd.edu

1 Introduction

Recently considerable attention has been given to the study of risk-sensitive stochasic control problems [3] [11] [16] [17] [19]. This type of problem was formulated first in [15], and in the operations research literature in [14]. One of the key results that has been explored in this study is the relationship between risk-sensitive stochastic control problems and game theory [2] [6] [10] [12] [18].

In this paper we are concerned with infinite horizon risk sensitive stochastic control problems with denumerable state space, discrete time parameter, and bounded cost function. The solution of this problem is based on the existence of a bounded solution to the corresponding dynamic programming equation, which turns out to be a nonlinear eigenvalue problem. In [12] a similar problem is studied for continuous variable systems modelled by differential equations. However, the approach we follow is technically different. For finite state space this problem was solved in [10] using the Perron-Frobenious Theorem and the policy iteration algorithm. Risk sensitive control problems on a finite horizon for hidden Markov models were treated in [9].

In this paper we determine sufficient conditions for the existence of a solution to the dynamic programming equation. Our results are based on the observation that the dynamic programming equation can be seen as the Isaacs equation of an ergodic cost stochastic dynamic game (see [1, 12]). Furthermore, we prove that these sufficient conditions are also necessary.

The remainder of the paper is organized as follows. In Section 2 we introduce the risk sensitive control problem and a verification theorem is presented. Section 3 discusses sufficient conditions for the existence of a bounded solution to the dynamic programming equation. Finally, in Section 4 we prove that some of the sufficient conditions introduced in Section 3 are also necessary.

2 Preliminaries

The model. The discrete-time Markov control model we will be dealing with is the following (see e.g. [1] [13]). Let (S, A, P, c) be a Markov control model where the state space S is a (nonempty) denumerable set endowed with the discrete topology, and A is a Borel space, called the action or control space. For every $x \in S$, $A(x)$ represents the set of admissible actions when the system is in state x. The set of admissible pairs is defined by $K = \{(x,a) : \ a \in A(x), x \in S\}$, and is considered as a topological subspace of $S \times A$. The transition law P is a stochastic kernel on S given K, and finally, $c : K \to \mathbb{R}$ is the (measurable) one-stage cost function.

Consider the history spaces $H_o = S$, and $H_t := \ K \times H_{t-1}$ if $t = 1, 2, \ldots$. An element h_t of H_t is a vector of the form $h_t = (x_o, a_o, \ldots, x_{t-1}, x_t)$, where $(x_n, a_n) \in K$ for $n = 0, \ldots, t-1$, and $x_t \in S$. An admissible control policy, or strategy, is a sequence $\pi = \{\pi_t\}$ of stochastic kernels π_t on A given H_t, satisfying the constraint $\pi_t(A(x_t)|h_t) = 1$, for all $h_t \in H_t, t \geq 0$. The set of policies is denoted by $\mathcal{P}$. Now, let $\mathbf{F}$ be the set of functions $f : S \to A$, such that $f(x) \in A(x)$ for all $x \in S$. A policy $\pi \in \mathcal{P}$ is stationary if there exist $f \in \mathbf{F}$ such that $\pi_t(f(x_t)|h_t) = 1$ for all $h_t \in H_t, t \geq 0$.

Let $(\Omega, \mathcal{F})$ be the measurable space that consists of the sample space $\Omega := (S \times A)^\infty$ and the corresponding product σ-algebra $\mathcal{F}$. Then, given any initial distribution p_o, for each policy $\pi \in \mathcal{P}$ a probability measure $P^\pi_{p_o}$ and a stochastic process $\{(x_t, a_t), t = 0, 1 \ldots\}$ are defined on $(\Omega, \mathcal{F})$ in a canonical way, where x_t and a_t denote the state and action at time t, respectively. When p_o is the Dirac measure concentrated at $x \in S$, we write P^π_x instead of $P^\pi_{p_o}$, and the corresponding expectation operator is denoted by E^π_x.

Throughout we assume the following, even when we do not mention it explicitly.

Assumption A.1.

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

(ii) The cost function c is nonnegative, continuous and bounded.

(iii) For all $x, y \in S$, the function $a \mapsto P(y|x,a)$ is continuous on $A(x)$.

Risk sensitive control problem. For $x \in S, \pi \in \mathcal{P}$, the cost functional (to be minimized) is the infinite horizon exponential growth criterion

$$J(x,\pi) := \limsup_{T\to\infty} \frac{1}{T}\ \gamma\ \log\ E^\pi_x\ \exp\ \{\frac{1}{\gamma}\ \sum_{t=0}^{T-1} c(x_t,a_t)\},$$

where $\gamma > 0$ is the risk factor. Then the risk sensitive optimal control problem is to find a policy $\pi^* \in \mathcal{P}$ that minimizes $J(x,\pi)$. The corresponding value function is defined by

$$\Lambda(x) := \inf_{\pi \in \mathcal{P}} J(x, \pi).$$

The main problem we are concerned with is to find sufficient conditions to ensure the existence of an optimal stationary policy.

Verification theorem.

Theorem 2.1. Suppose that there exist a number λ and a bounded function $W : S \to \mathbb{R}$ such that

$$e^{\lambda + W(x)} = \min_{a \in A(x)} \{ e^{\frac{1}{\gamma} c(x,a)} \int e^{W(y)} P(dy|x, a) \} \text{ for all } x \in S. \tag{2.1}$$

Then, for each $x \in S$,

$$\lambda\gamma \leq J(x, \pi) \text{ for all } \pi \in \mathcal{P}.$$

Further, if f^* is a stationary policy, with $f^*(x)$ achieving the minimum on the r.h.s. of (2.1) for each $x \in S$, then f^* is optimal, and

$$\lambda = \lim_{T \to \infty} \frac{1}{T} \log E_x^{f^*} \exp \{ \frac{1}{\gamma} \sum_{t=0}^{T-1} c(x_t, a_t) \}.$$

Proof. Let $x \in S$, and $\pi \in \mathcal{P}$. Then, from (2.1)

$$\begin{aligned} E_x^\pi e^{\frac{1}{\gamma} \sum\limits_{t=0}^{T-1} c(x_t, a_t)} &\geq E_x^\pi \Pi_{t=0}^{T-1} \left[e^\lambda \cdot \frac{e^{W(x_t)}}{P(e^W)(x_t, a_t)} \right] \\ &= e^{\lambda T} E_x^\pi \left[\Pi_{t=0}^{T-1} \frac{e^{W(x_t)}}{P(e^W)(x_t, a_t)} \right], \end{aligned} \tag{2.2}$$

where $P(e^W)(x, a) := \int e^{W(y)} P(dy|x, a)$. Using standard properties of conditional expectations and the Markov property, we have that

$$E_x^\pi \left[\Pi_{t=0}^{T-1} \frac{e^{W(x_t)}}{P(e^W)(x_t, a_t)} \cdot P(e^W)(x_{T-1}, a_{T-1}) \right] = e^{W(x)}. \tag{2.3}$$

Then, since we are assuming that W is bounded, (2.3) implies the existence of suitable positive constants M_1 and M_2 such that

$$
\begin{aligned}
M_1 \;&\leq\; \frac{e^{W(x)}}{\sup\limits_{x\in S_a\in A(x)} P(e^W)(x,a)} \leq E_x^{\pi}\left[\Pi_{t=0}^{T-1}\frac{e^{W(x_t)}}{P(e^W)(x_t,a_t)}\right] \\
&\leq\; \frac{e^{W(x)}}{\inf\limits_{x\in S_a\in A(x)} P(e^W)(x,a)} \leq M_2. \qquad (2.4)
\end{aligned}
$$

Therefore, from (2.2) and (2.4) we get

$$\lambda\gamma \leq J(x,\Pi).$$

The rest of the proof follows immediately replacing the inequality sign in (2.2) by equality, and repeating the above arguments. ■

3 Dynamic programming equation

In this section we turn to the question of existence of a solution to the dynamic programming equation (2.1).

The main result of this paper is the following:

Theorem 3.1. For each $e \in S$ and $f \in \mathbf{F}$ define

$$\tau_e := \min\{t > 0 : x_t = e\}. \qquad (3.1)$$

If there exists $e \in S$ and $C > 0$ such that

$$E_x^f(\tau_e) < C \qquad (3.2)$$

for all $f \in \mathbf{F}$ and $x \in S$, then there exists a solution (λ, W) to the dynamic programming equation (2.1), with W bounded.

Remark 3.2. We anticipate that Theorem 3.1 can be generalized to weaker conditions as in [1, p. 302], and active research in this direction is presently in progress. Many sufficient conditions for (3.2) are well known; see [1] and references therein.

Before we give the proof of Theorem 3.1 we will introduce some preliminary results. Let $P(S)$ be the set of probability vectors on S, i.e.

$$P(S) = \{\mu = (\mu^0, \mu^1, \dots) : \mu^i \geq 0, \sum_{i=0}^{\infty} \mu^i = 1\}.$$

Let us fix $\nu \in P(S)$. We define the relative entropy function $I(\cdot||\nu) : P(S) \to \mathbb{R} \cup \{+\infty\}$ by

$$I(\mu||\nu) = \begin{cases} \sum\limits_{x \in S} log(r(x))\mu(x) & \text{if } \mu << \nu \\ +\infty & \text{otherwise} \end{cases}$$

where

$$r(x) = \begin{cases} \frac{\mu(x)}{\nu(x)} & \text{if } \nu(x) \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$

The next lemma establishes, using a Legendre-type transformation, the duality relationship between the relative entropy function and the logarithmic moment generating function. For its proof we refer to [7, Proposition II.4.2].

Lemma 3.3. Let ψ be a bounded function defined on S, and let $\nu \in P(S)$. Then,

$$\log \int e^{\psi(y)} \nu(dy) = \sup_{\mu \in P(S)} \{\int \psi(y)\mu(dy) - I(\mu||\nu)\},$$

and the supremum is attained at the unique probability measure μ^* defined by

$$\mu^*(x) = \frac{e^{\psi(x)}}{\int e^{\psi} d\nu} \nu(x), \quad x \in S.$$

Therefore, using Lemma 3.3, we rewrite equation (2.1) as

$$\lambda + W(x) = \min_{a \in A(x)} \sup_{\mu \in P(S)} \{\int W d\mu + \frac{1}{\gamma} c(x,a) - I(\mu||P(\cdot|x,a))\}. \tag{3.3}$$

This equation corresponds to the Isaacs equation associated with a stochastic dynamic game with average cost per unit time criterion (see [10]).

The technique we will follow to prove Theorem 3.1 is the standard vanishing discount approach (see, e.g. [1] and the references therein). We consider the corresponding infinite horizon discounted cost stochastic dynamic game. Let W_β be the upper value function of this game. Then, once we find a uniform bound for a "differential" discounted value function, i.e. $h_\beta(x) := W_\beta(x) - W_\beta(e)$, with e as in (3.2), the theorem follows by letting $\beta \to 1$.

First we introduce the infinite horizon discounted cost dynamic game.

Stochastic dynamic game. Let S be the state space, A be the control set for Player 1 (minimizer), and $P(S)$ be the control set for Player 2 (maximizer). The reward function is $(x, a, \mu) \mapsto \frac{1}{\gamma}c(x, a) - I(\mu||P(\cdot|x, a))$, with $(x, a, \mu) \in K \times P(S)$.

The evolution of the system is as follows (c.f. [10] [12]). At each time $t \in \{0, 1, \ldots\}$ the state of the system is observed, say $x_t = x \in S$. Then, a control $a_t \in A(x)$ is chosen for Player 1, and $\mu_t \in P(S)$ is chosen for Player 2. Then, a reward $\frac{1}{\gamma}c(x_t, a_t) - I(\mu_t||P(\cdot|x_t, a_t))$ is earned, and the state of the system moves to the state x_{t+1} according to the probability distribution μ_t.

Strategies. For each $t \geq 0$, let N_t and K_t be the set of feasible histories up to time t for Player 1 and Player 2, respectively. That is, $N_0 = S$ and $N_t = (S \times P(S))^t \times S$, while $K_0 = K$ and $K_t = K^t \times K$. Generic elements of N_t and K_t are vectors of the form $n_t = (x_0, \mu_0, \ldots, x_{t-1}, \mu_{t-1}, x_t)$ and $K_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t, a_t)$, respectively. A non-randomized strategy for Player 1 is a sequence $\vec{f} = \{f_t\}$ of functions f_t from N_t to A, such that $f_t(n_t) \in A(x_t)$ for all $n_t \in N_t$. We say that $\vec{f}$ is stationary if, for all $t \geq 0$, f_t depends only on the current state x_t, and $f_t = f \in \mathbf{F}$ is independent of t. A non-randomized strategy for Player 2 is a sequence $\vec{\xi} = \{\xi_t\}$ of functions ξ_t from K_t to $P(S)$. Analogously, $\vec{\xi}$ is stationary if, for all $t \geq 0, \xi_t = \xi : K \mapsto P(S)$.

Given the initial state $x \in S$, and the strategies $\vec{f}, \vec{\xi}$ being used, the corresponding state process, x_t, is a stochastic process defined on the canonical probability space $(S^\infty, \mathcal{G}, P_x^{\vec{f},\vec{\xi}})$, where $P_x^{\vec{f},\vec{\xi}}$ is uniquely determined. We denote by $E_x^{\vec{f},\vec{\xi}}$ the corresponding expectation operator.

Equation (3.3) corresponds to the dynamic progamming (Isaacs) equation of the stochastic dynamic game described above with average cost optimality criterion, defined for each $x \in S, \vec{f}, \vec{\xi}$ as

$$\Lambda(x, \vec{f}, \vec{\xi}) := \limsup_{T \to \infty} E^{\vec{f}, \vec{\xi}_x} \frac{1}{T} \sum_{t=0}^{T-1} \left[\frac{1}{\gamma} c(x_t, a_t) - I(\xi_t || P(\cdot | x_t, a_t))\right].$$

Then, in order to prove Theorem 3.1, we will study a sequence of corresponding discounted games. Thus, given $x \in S, \vec{f}, \vec{\xi}$, define the cost functional

$$J_\beta(x, \vec{f}, \vec{\xi}) = E_x^{\vec{f}, \vec{\xi}} \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{\gamma} c(x_t, a_t) - I(\xi_t || P(\cdot | x_t, a_t))\right], \tag{3.4}$$

where $\beta \in (0, 1)$ is the discount factor.

Definition 3.4. When there exist a pair of strategies $(\vec{f^*}, \vec{\xi^*})$ such that

$$J_\beta(x, \vec{f^*}, \vec{\xi}) \leq J_\beta(x, \vec{f^*}, \vec{\xi^*}) \leq J_\beta(x, \vec{f}, \vec{\xi^*})$$

for all $\vec{f}, \vec{\xi}$, the value $W_\beta(x) = J_\beta(x, \vec{f^*}, \vec{\xi^*})$ is called the value of the game, and $(\vec{f^*}, \vec{\xi^*})$ are referred to as *optimal strategies.*

Lemma 3.5. There is a unique bounded solution to the Isaacs equation

$$W_\beta(x) = \min_{a \in A(x)} \sup_{\mu \in P(S)} \{\int \beta W_\beta d\mu + \frac{1}{\gamma} c(x, a) - I(\mu || P(\cdot | x, a))\}, \tag{3.5}$$

and it is the value function of the discounted cost stochastic dynamic game. Moreover, the stationary strategies f^* and ξ^*, with

$$f^*(x) \in \arg \min_{a \in A(x)} \{e^{\frac{1}{\gamma} c(x,a)} \int e^{\beta W_\beta(y)} P(dy | x, a)\}$$

and

$$\xi^*[x, a](x'') = \frac{e^{\beta W_\beta(x'')} P(x'' | x, a)}{\int e^{\beta W_\beta(y)} P(dy | x, a)} \tag{3.6}$$

are optimal.

Proof. Let $B(S)$ be the vector space of real-valued bounded functions defined on S endowed with the supremum norm, i.e. for $\psi \in B(S), ||\psi|| = \sup_{x \in S} |\psi(x)|$. Now, define the operator $L : B(S) \to B(S)$ by

$$L\psi(x) := \min_{a\in A(x)} \sup_{\mu\in P(\mu)} \{\int \beta\psi d\mu + \frac{1}{\gamma}c(x,a) - I(\mu| - P(\cdot(x,a))\}.$$

Then, the first part of the lemma follows noting that L is a contraction operator and the Fixed Point Theorem. The rest of the proof is a straight-forward application of standard dynamic programming arguments and Lemma 3.3. ■

Remark 3.6. Note that, by Lemma 3.3, equation (3.5) can be rewritten as

$$e^{W_\beta(x)} = \min_{a\in A(x)} \{e^{\frac{1}{\gamma}c(x,a)} \int e^{\beta W_\beta(y)} P(dy|x,a)\}. \tag{3.7}$$

This optimality equation has been studied by Chung and Sobel [5] (see also the references therein), in the context of risk-sensitive discounted Markov decision processes.

Now, following a standard approach, we define $h_\beta(x) := W_\beta(x) - W_\beta(e)$, with e as in (3.2), and write (3.7) as

$$e^{(1-\beta)W_\beta(e)} \cdot e^{h_\beta(x)} = \min_{a\in A(x)} e^{\frac{1}{\gamma}c(x,a)} \int e^{\beta h_\beta(y)} P(dy|x,a)\}. \tag{3.8}$$

Proof of Theorem 3.1. We will prove first that $(1-\beta)W_\beta(e)$ and h_β are uniformly bounded. Note that from (3.3) and the definition of W_β we have

$$0 \le (1-\beta)W_\beta(x) \le \frac{1}{\gamma}||c|| \text{ for all } x \in S, \tag{3.9}$$

where $||c||$ stands for the supremum norm of c. We next consider h_β. Let $\beta \in (0,1)$, and let f^*, ξ^* be the optimal stationary policies for Player 1 and Player 2 defined in Lemma 3.5. Then, using the fact that the stochastic kernel of the Markov process x_t, say P^*, defined by the strategies f^*, ξ^* has the form

$$P^*(x''|x) = \frac{e^{\beta W_\beta(x'')} P(x''|x, f^*(x))}{\int e^{\beta W_\beta(y)} P(dy|x, f^*(x))},$$

(see (3.6)), we have

$$W_\beta(x) = E_x^{f^*,\xi^*} \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{\gamma}c(x_t,a_t) - I(\xi_t||P(\cdot|x_t,a_t))\right]$$

$$= E_x^{f^*} \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{\gamma} c(x_t, a_t) - log\left(\frac{e^{\beta W_\beta(x_{t+1})}}{P(e^{\beta W_\beta}(x_t, a_t)}\right) \frac{e^{\beta W_\beta(x_{t+1})}}{P(e^{\beta W_\beta}(x_t, a_t)}\right]$$

$$\cdot \left[\frac{e^{\beta W_\beta(x_t)}}{P(e^{\beta W_\beta})(x_{t-1}, a_{t-1})} \cdots \frac{e^{\beta W_\beta(x_1)}}{P(e^{\beta W_\beta})(x_0, a_0)}\right]$$

$$= E_x^{f^*} \sum_{t=0}^{\infty} \beta^t r(x_0, a_0, \ldots, x_t, a_t, x_{t+1}), \tag{3.10}$$

with $P(e^{\beta W_\beta})(x,a) := \int e^{\beta W_\beta(y)} P(dy|x,a)$, and r is defined implicitly.
Then, by (3.1) and (3.10)

$$\begin{aligned} W_\beta(x) &= E_x^{f^*} \sum_{t=0}^{\tau-1} \beta^t r(x_0, a_0, \ldots, x_t, a_t, x_{t+1}) + E_e^{f*} \sum_{t=\tau}^{\infty} r(x_\tau, a_\tau, \ldots, x_t, a_t, x_{t+1}) \\ &\leq C\frac{1}{\gamma} \cdot ||c|| + W_\beta(e), \end{aligned}$$

therefore,

$$W_\beta(x) - W_\beta(e) \leq C \cdot \frac{1}{\gamma} ||c||. \tag{3.11}$$

Also, from (3.7) and Jensen's inequality, and using the fact that c is nonnegative, it follows that

$$W_\beta(x) \geq \beta E_x^{f^*} W_\beta(x_1).$$

Actually, we can prove, using the same arguments, by induction that for every $T > 0$

$$W_\beta(x) \geq E_x^{f^*} \beta^T W_\beta(x_T),$$

which, in particular, implies that

$$\begin{aligned} W_\beta(x) &\geq E_x^{f^*} \beta^\tau W_\beta(x_\tau) \\ &= W_\beta(e) E_x^{f^*} \beta^\tau \geq W_\beta(e) \beta^C, \end{aligned}$$

where the last inequality is due to Jensen's inequality and (3.2).
Therefore,

$$\begin{aligned} W_\beta(e) - W_\beta(x) &\leq (1-\beta^C)W_\beta(e) \\ &\leq (1-\beta^C)\cdot\frac{\frac{1}{\gamma}||c||}{1-\beta} \leq C\cdot\frac{1}{\gamma}||c||. \end{aligned} \tag{3.12}$$

Thus, from (3.11)-(3.12) we get

$$|h_\beta(x)| \leq C\cdot\frac{1}{\gamma}||c|| \quad \text{for all } x\in S \tag{3.13}$$

Let $\beta_n \uparrow 1$ be given. Then, (3.9) and (3.13) imply that, by a suitable diagnolization, we may pick a subsequence $\{\beta_n\}$ (denoting it again by $\{\beta_n\}$) along which $h_{\beta_n}(x), x\in S$, and $(1-\beta)W_\beta(e)$ converge to some limits $W(x)$ and λ, respectively. Thus, the theorem follows from (3.8) and an application of the Dominated Convergence Theorem. ∎

4 Necessary conditions

In the previous section we proved that sufficient conditions to ensure the existence of a bounded solution to the dynamic programming equation (2.1) are the following.
(C1) Suppose that the unique bounded solution W_β of the equation (3.5) satisfies:

(i) $x\mapsto(1-\beta)W_\beta(x)$ is uniformly bounded.

(ii) $h_\beta(x) := W_\beta(x) - W_\beta(e)$ is uniformly bounded, with $e\in S$ being arbitrary, but fixed.

The purpose of this section is to prove that (C1) is actually a necessary condition for the existence of a bounded solution to the equation (2.1). See [4], [8], where necessary conditions for the existence of a bounded solution of the risk-neutral average cost optimality equation are given. Notice first that, if (λ, W) is a solution of (2.1), then, for any constant $k, (\lambda, W+k)$ also satisfies (2.1).

Theorem 4.1. Assume that there exist a number λ and a bounded function $W : S \to \mathbb{R}$ satisfying (2.1). Then,

$$W_1(x) \le \beta W_\beta(x) - \frac{\beta\lambda}{1-\beta} \le W_2(x) \quad \text{for al } x \in S, \tag{4.1}$$

where W_1 and W_2 are solutions of (2.1) such that $\sup_{x\in S} W_1(x) = \inf_{x\in S} = W_x(x) = 0$. In particular,

$$\lim_{\beta\to 1}(1-\beta)W_\beta(x) = \lambda \quad \text{for all } x \in S,$$

and

$$|W_\beta(x) - W_\beta(y)| \le \frac{1}{\beta}\left[||W_1|| + ||W_2||\right].$$

Proof. Given a bounded function $\psi : S \to \mathbb{R}$, we define the operator

$$T\psi(x) = \min_{a\in A(x)}\{e^{\frac{1}{\gamma}c(x,a)}\int \psi(y)P(dy|x,a)\}.$$

Then, we can rewrite equations (2.1) and (3.7) as

$$e^{\lambda + W(x)} = Te^{W(x)}$$

$$e^{W_\beta(x)} = Te^{\beta W_\beta(x)}. \tag{4.2}$$

Let $\bar{W}(x) = \beta W_\beta(x) - \frac{\beta\lambda}{1-\beta}$. Then, from (4.2)

$$\begin{aligned} T(e^{\bar{W}})(x) &= \bar{e}^{\frac{\beta\lambda}{1-\beta} + W_\beta(x)} \\ &= e^{\lambda}(e^{\bar{W}(x)})^{\frac{1}{\beta}} \\ &= \frac{T(e^{W_1})(x)}{e^{W_1(x)}}(e^{\bar{W}(x)})^{\frac{1}{\beta}}. \end{aligned}$$

Therefore,

$$\frac{e^{W_1(x)}}{e^{\bar{W}(x)}} \cdot \frac{T(e^{\bar{W}})(x)}{T(e^{W_1})(x)} = \left(e^{\bar{W}(x)}\right)^{\frac{1-\beta}{\beta}} \tag{4.3}$$

Suppose that $1 > \inf_{x \in S} \left[\frac{e^{\bar{W}(x)}}{e^{W_1(x)}} \right] \equiv \rho$. Let $\epsilon > 0$ and $x_0 \in S$ such that

$$1 > \rho \geq \frac{e^{\bar{W}(x_0)}}{e^{W_1(x_0)}} - \epsilon > 0. \tag{4.4}$$

Then, from (4.4) and the normalization condition of W_1, it follows

$$\left[e^{\bar{W}(x_0)} \right]^{\frac{1-\beta}{\beta}} \leq (\rho + \epsilon)^{\frac{1-\beta}{\beta}}. \tag{4.5}$$

Also,

$$\begin{aligned} \frac{e^{W_1(x_0)}}{e^{\bar{W}(x_0)}} \cdot \frac{T(e^{\bar{W}})(x_0)}{T(e^{W_1})(x_0)} &\geq \rho \cdot e^{W_1(x_0) - \bar{W}(x_0)} \\ &\geq \frac{\rho}{\rho + \epsilon} \end{aligned} \tag{4.6}$$

Thus, from (4.3)-(4.6) we have

$$\frac{\rho}{\rho + \epsilon} \leq (\rho + \epsilon)^{\frac{1-\beta}{\beta}}.$$

Since ϵ was chosen arbitrarily small, we get

$$1 \leq \rho^{\frac{1-\beta}{\beta}},$$

which is a contradiction. Then, the left hand side of (4.1) holds. The proof of the r.h.s. follows in a similar way and we omit it. Employing (4.1) yields the rest of the theorem. ∎

References

[**1**] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus, Discrete-time controlled Markov processes with average cost criterion: a survey, SIAM J. Control and Optim. (1993) 31, 282-344.

[**2**] J. S. Baras and M. R. James, Robust and risk-sensitive output feedback control for finite state machines and hidden Markov models, submitted to J. Math. Systems, Estimation and Control.

[**3**] A. Bensoussan and J. H. Van Schuppen, Optimal control of partially observable stochastic systems with exponential of integral performance index, SIAM J. Control and Optim. (1985) 23, 599-613.

[**4**] R. Cavazos-Cadena, Necessary conditions for the optimality equation in average reward Markov decision processes, Applied Math. and Optim. (1989) 19, 97-112.

[**5**] K.-J. Chung and M. J. Sobel, Discounted MDP's: Distribution functions and exponential utility maximization, SIAM J. Control and Optim. (1987) 25, 49-62.

[**6**] P. Dai Pra, L. Meneghini and W. J. Runggaldier, Some connections between stochastic control and dynamic games, preprint.

[**7**] P. Dupuis and R. S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations, forthcoming, to be published by John Wiley & Sons.

[**8**] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus, Remarks on the existence of solutions to the average cost optimality equation in Markov decision processes, Systems Control Lett. (1990) 15, 425-432.

[**9**] E. Fernández-Gaucherand and S. I. Marcus, Risk-sensitive optimal control of hidden Markov models: a case study, Proc. 33rd CDC (1994), 1657-1662.

[**10**] W. H. Fleming and D. Hernández-Hernández, Risk sensitive control of finite state machines on an infinite horizon, preprint.

[11] W. H. Fleming and W. M. McEneaney, Risk-sensitive control and differential games, Springer Lecture Notes in Control and Info. Sci. No. 184, 1992, 185-197.

[12] W. H. Fleming and W. M. McEneaney, Risk-sensitive control on an infinite horizon, SIAM J. Control and Optim. (1995) 33,1881-1915.

[13] O. Hernández-Lerma, Adaptive Markov Control Processes, Springer-Verlag, New York, 1989.

[14] R. A. Howard and J. E. Matheson, Risk-sensitive Markov decision processes, Management Sci. (1972) 18, 356-369.

[15] D. H. Jacobson, Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games, IEEE Trans. Aut. Control (1973) 18, 124-131.

[16] M. R. James, Asymptotic analysis of nonlinear stochastic risk-sensitive control and differential games, Math. of Control, Signals and Systems (1992) 5(4), 401-417.

[17] M. R. James, J. S. Baras and R. J. Elliott, Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems, IEEE Trans. Aut. Control (1994) 39, 780-792.

[18] T. Runolfsson, The equivalence between infinite horizon control of stochastic systems with exponential-of-integral performance index and stochastic differential games, IEEE Trans. Aut. Control (1994) 39, 1551-1563.

[19] P. Whittle, Risk-Sensitive Optimal Control, John Wiley & Sons, New York, 1990.