ON DUALITY OF DISCRETE TIME STOCHASTIC CONTROL PROCESSES

Kimura, Yutaka
Department of Mathematical Sciences, Graduate School of Science and Technology, Niigata University

Tanaka, Kensuke
Department of Mathematics, Faculty of Science, Niigata Univerisity

https://doi.org/10.5109/13464

出版情報：Bulletin of informatics and cybernetics. 29 (1), pp.91-103, 1997-03. Research Association of Statistical Sciences
バージョン：
権利関係：

Bulletin of Informatics and Cybernetics, Vol. 29, No. 1, 1997

ON DUALITY OF DISCRETE TIME STOCHASTIC CONTROL PROCESSES

By

Yutaka KIMURA* and Kensuke TANAKA†

Abstract

In this paper, we investigate dynamic programming models with a discrete time and an infinite horizon. The main purpose is to seek an optimal value and an optimal policy under various conditions. To do this, we introduce a modified form of the dynamic model which we call a duality of the dynamic one. Then we show that an optimal value of the original model is equal to the one of the dual model, and that there exists an optimal dual policy for the dual model. Further, in view of the dual model we show that there exists an optimal policy for the original model.

1. Introduction

Dynamic programming problems with a discrete time and an infinite horizon have been investigated by many authors. Some of the main earlier works in this area were done by Blackwell (1965) and Strauch (1966). Further, Dynkin and Yushkevich (1979) gave extensive accounts of dynamic programming with a discrete time parameter. In many cases of these researches, the concept of optimal policy is given and, then one of our purposes is to seek an optimal value and an optimal policy under the various conditions. In order to show the existence of an optimal policy, we will study it under some condition, for example, that the action space is compact. See Balder (1989) and Schäl (1975) in detail. Thus, an optimal policy may not exist if compactness is weakened. In this case, we study the properties of ε-optimal policies as in Tanaka, Hoshino, and Kuroiwa (1995).

On the other hand, the various dual models have been studied as another approaches for dynamic model. See Iwamoto (1983), (1984), and Tanaka (1995) in details.

In optimization theory, dual optimization problems on a dual space are introduced. Using the properties of conjugate functions and hyperplanes on a vector space, the relations between the original optimization problems and their duals have been actively discussed on the basis of convex analysis. See Aubin (1982), (1993), Luenberger (1969), and Rockafellar (1966) in details.

* Department of Mathematical Science, Graduate School of Science and Technology, Niigata University, 950-21, Niigata, Japan

† Department of Mathematics, Faculty of Science, Niigata University, 950-21, Niigata, Japan

In this paper, we introduce a modified form of the dynamic model, which we shall call a "*dual*" dynamic model. A reward function in the dual model is a conjugate function for loss function in the original model. Thus, the reward function has good properties which are convex and w^*-lower semicontinuous. Here, using these properties, we show that the optimal value of the original problem is equal to the optimal one in the dual model. Further, we show that there exists an optimal dual policy in the dual model, which we call a weak optimal policy for the original model. To do this, Fenchel's inequality and the Fenchel duality theorem in convex analysis (see for example Aubin (1993)) play very important roles.

This paper is organized in the following way. In Section 2, we formulate a basic minimization problem relative to the dynamic programming model and give the definitions of optimal value and optimal policy. In Section 3, we give some basic results for the dynamic model. In Section 4, we introduce the concept of a dual space, and, then give a dual form of the original model. Further, using Fenchel duality theorem, we show that there exists an optimal dual policy in the dual model, and we discuss the relation between the original model and the dual one.

2. Formulation of a dynamic programming model

A dynamic programming model is specified by a set of six elements

$$(S,\ A,\ A(S),\ q,\ r,\ \beta), \tag{2.1}$$

where we assume :

(i) S is a Polish space (a complete separable metric space). The Borel measurable space $(S, \mathcal{B}(S))$ is called the *state space* of the dynamic model, where $\mathcal{B}(S)$ is the Borel field of S.

(ii) A is a real Banach space and $(A, \mathcal{B}(A))$ is called the *action space*, where $\mathcal{B}(A)$ is the Borel field of A.

(iii) For each $s \in S$, $A(s) \subset A$ is the set of all admissible actions on the state s, where $A(S) = \bigcup_{s \in S} A(s)$. We assume that $\mathrm{Gr}\, A = \{(s,a) | s \in S, a \in A(s)\}$ is Borel measurable in SA, where SA denotes the Cartesian product of sets S and A.

(iv) q is a *transition probability measure* on S given SA, that is, $q(\cdot|s,a)$ is a probability measure on $(S, \mathcal{B}(S))$ for each $(s,a) \in \mathrm{Gr}\, A$ and $q(\Gamma|\cdot,\cdot)$ is a Borel measurable function on $\mathrm{Gr}\, A$ for each Borel subset $\Gamma \in \mathcal{B}(S)$. The law of motion of the dynamic model is given by q.

(v) $r : \mathrm{Gr}\, A \to \mathbb{R}_+$ is an *one-stage loss function*, which is Borel measurable, where $\mathbb{R}_+ = \{r \in \mathbb{R} | r \geq 0\}$.

(vi) $\beta \in (0,1)$ is a *discount factor* .

In order to consider the dual model, it will be needed that the loss function is bounded from below. Thus, we give condition (v). Furthermore, in the specification, we should note that the permissible set of actions $A(s)$ depends on the state $s \in S$ and $q(\cdot|s,a)$ is independent of the time parameter.

Then, a policy π for the model is defined as an infinite sequence $\pi = (f_0, f_1, \cdots, f_k, \cdots)$, each component f_k of which is a Borel measurable mapping from S to A such that $f_k(s_k) \in A(s_k)$ for every $s_k \in S$, where s_k denotes state on the k-th stage. Since each component f_k of a policy π in the model is parametrized only by s_k, π is said to be a *Markov policy*. We denote by Π the set of all Markov policies. If π is a Markov policy of the form $\pi = (f, f, \cdots)$, it is said to be *stationary* and denoted by f^∞.

The model is interpreted as follows. If a policy $\pi = (f_0, f_1, \cdots, f_k, \cdots)$ is employed, at the successive k-stages, $k = 0, 1, 2, \cdots$, we observe $s_k \in S$, and then we choose an action $a_k \in A(s_k)$ according to k-th component f_k of π, that is, $a_k = f_k(s_k)$. As a result, we will incur a loss $r(s_k, a_k)$. Then, the dynamic model moves to a new state $s_{k+1} \in S$ according to the stochastic kernel $q(\cdot|s_k, a_k)$, and the process is analogously developed from s_{k+1}.

In the subsequent discussion, we will often use S_k and A_k for $k = 0, 1, 2, \cdots$, as copies of S and A, respectively. Given an initial distribution μ on S and a policy $\pi = (f_0, f_1, f_2, \cdots)$ together with the stochastic kernel q, there is a sequence of unique probability measures $P_k^{\pi,\mu}$ on $S_0 S_1 \cdots S_k$ $(k = 0, 1, \cdots)$. In addition, there exists a unique probability measure $P^{\pi,\mu}$ on $S_0 S_1 \cdots$ such that, for each k, marginal measure of $P^{\pi,\mu}$ on $S_0 S_1 \cdots S_k$ is $P_k^{\pi,\mu}$.

Then, if we assume that a policy $\pi = (f_0, f_1, \cdots)$ is used under the discount factor β, the *expected loss* on the k-stage is given by

$$\begin{aligned} E_\pi^\mu [r(s_k, f_k(s_k))] &= \int_{S_0 S_1 \cdots S_k} r(s_k, f_k(s_k)) P_k^{\pi,\mu}(d(s_0, s_1 \cdots, s_k)) \\ &= \int_{S_0} \mu(ds_0) \int_{S_1} q(ds_1|s_0, f_0(s_0)) \cdots \\ &\quad \int_{S_k} r(s_k, f_k(s_k)) q(ds_k|s_{k-1}, f_{k-1}(s_{k-1})), \end{aligned}$$

and the *total expected discounted loss* is given by

$$I_\pi(\mu) = \sum_{k=0}^{\infty} \beta^k E_\pi^\mu [r(s_k, f_k(s_k))]. \tag{2.2}$$

Under condition (v), the expectation and summation in (2.2) can be interchanged to obtain

$$I_\pi(\mu) = E_\pi^\mu \left[\sum_{k=0}^{\infty} \beta^k r(s_k, f_k(s_k)) \right].$$

We often write $I_f(\mu)$ instead of $I_{f^\infty}(\mu)$ for a stationary policy $f^\infty = (f, f, \cdots)$. Further, if the initial distribution μ assigns mass one to a point $s \in S$, we use $I_\pi(s)$ instead of $I_\pi(\mu)$.

Then, we consider a basic minimization problem **(MP)** for the dynamic model :

$$\textbf{(MP)} \qquad \text{minimize} \quad I_\pi(s) \qquad \text{subject to} \quad \pi \in \Pi.$$

DEFINITION 2.1. The optimal discounted loss $I(s)$ is defined by

$$I(s) = \inf_{\pi \in \Pi} I_\pi(s), \qquad s \in S, \tag{2.3}$$

and this is said to be *the optimal value for the initial state* s.

DEFINITION 2.2. If $I(s) = I_{\bar{\pi}}(s)$, the policy $\bar{\pi} \in \Pi$ is said to be *optimal for the initial state* s. If $\bar{\pi}$ is optimal for every initial state s, it is said to be *optimal.*

In this paper, we will study the problem **(MP)** under the assumption such that the optimal value is finite for all initial states, that is, $I(s) \in \mathbb{R}_+$ for all $s \in S$.

3. Basic results for the dynamic programming problem

In order to state some basic results for the model with condition (v), let $B(S)$ be the space of all real-valued measurable and non-negative functions on S. We give an operator T_f on $B(S)$ defined by

$$T_f v(s) = r(s, f(s)) + \beta E_f[v|s], \qquad v \in B(S),\ s \in S, \tag{3.1}$$

where

$$E_f[v|s] = \int_S v(s')q(ds'|s, f(s))$$

and, furthermore, give an operator T on $B(S)$ defined by

$$Tv(s) = \inf_{f \in M} T_f v(s), \qquad v \in B(S),\ s \in S, \tag{3.2}$$

where M is the set of all measurable mappings f from S to A such that, for each $s \in S$, $f(s) \in A(s)$.

We shall now state some results for problem (MP) on Π. For a Markov policy $\pi = (f_0, f_1, \cdots)$ and an initial distribution μ_s which assigns mass one to a state $s \in S$, the *expected discounted loss up to the t-stage for* μ_s is written by

$$I_\pi^t(s) = \sum_{k=0}^{t} \beta^k E_\pi^{\mu_s}[r(s_k, f_k(s_k))] = E_\pi\left[\sum_{k=0}^{t} \beta^k r(s_k, f_k(s_k)) \,\middle|\, s_0 = s\right],$$

and, further, the *total expected discounted loss for* μ_s is written by

$$I_\pi(s) = \sum_{k=0}^{\infty} \beta^k E_\pi^{\mu_s}[r(s_k, f_k(s_k))] = E_\pi\left[\sum_{k=0}^{\infty} \beta^k r(s_k, f_k(s_k)) \,\middle|\, s_0 = s\right].$$

THEOREM 3.1. *Suppose that* $v_0 \equiv 0$. *Then for any Markov policy* $\pi = (f_0, f_1, \cdots)$,

$$I_\pi^t(s) = T_{f_0} T_{f_1} \cdots T_{f_t} v_0(s), \tag{3.3}$$

and

$$I_\pi(s) = \lim_{t \to \infty} T_{f_0} T_{f_1} \cdots T_{f_t} v_0(s). \tag{3.4}$$

THEOREM 3.2. *Suppose that* $v_0 \equiv 0$. *Then, we have the following :*

(i) $\inf_{\pi \in \Pi} I_\pi^t(s) = T^{t+1} v_0(s)$ *for all* $t \geq 0$

(ii) $I(s) = TI(s)$ *(dynamic programming equation)*

where $I(s)$ *is the optimal value for an initial state* $s \in S$ *and* Π *is the set of all Markov policies.*

The proofs of these theorems are shown in Bertsekas and Shreve (1978), and Dynkin and Yushkevich (1979) in detail.

4. Dual form of the dynamic model and its properties

In addition to the results in Section 3, we want to find an optimal policy for the dynamic model. In order to give an approach for the dynamic model, we introduce a dual form and show that the optimal value of the dual form is equal to I in Theorem 3.2. Further, we shall show that there exists an optimal policy in the dual form and discuss the relations between the original model and the dual one.

Now, we use the notation for the domain of $E.[v|s]$ for each $s \in S$ and $v \in B(S)$,

$$B_s(v) = \{b \in A(s) | E_b[v|s] < +\infty\} \subset A(s).$$

ASSUMPTION 4.1. There exists an $f \in M$ such that, for all $s \in S$ and $v \in B(S)$,

$$f(s) \in B_s(v).$$

Under this assumption, there exists an $f \in M$ such that $T_f : B(S) \to B(S)$. Thus, we have $Tv(s) < +\infty$.

To formulate a dual dynamic model, we give the preliminaries. Let A^* be the dual space of A. The norm of the dual space is given in the usual way, that is,

$$\|a^*\|_* = \sup \left\{ \langle a\,,\, a^* \rangle \;\middle|\; \|a\| \leq 1, a \in A \right\}, \qquad a^* \in A^*,$$

where $\langle a\,,\, a^* \rangle$ denotes the duality pairing of a and a^*. Thus, from the property of the dual space, it follows that A^* is Banach space. See Chapter 5 in Luenberger (1969) for details.

For each $s \in S$, the function $r^*(s, \cdot)$ from A^* to $\overline{\mathbb{R}}$ defined by

$$r^*(s, a^*) = \sup_{a \in A(s)} \{\langle a\,,\, a^* \rangle - r(s, a)\}$$

is called the (Fenchel) conjugate function of $r(s, \cdot)$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. Similarly, for each $s \in S$ and $v \in B(S)$, the function $E_{\cdot}^{*}[v|s]$ from A^* to $\overline{\mathbb{R}}$ defined by

$$E_{b^*}^{*}[v|s] = \sup_{a \in B_s(v)} \{\langle a, b^* \rangle - E_a[v|s]\} \tag{4.1}$$

is the conjugate function of $E_{\cdot}[v|s]$.

Consequently, from (4.1), we obtain

$$\begin{aligned}(\beta E_{\cdot}[v|s])^* &= \sup_{a \in B_s(v)} \{\langle a, a^* \rangle - \beta E_a[v|s]\} \\ &= \beta \sup_{a \in B_s(v)} \left\{ \left\langle a, \frac{a^*}{\beta} \right\rangle - E_a[v|s] \right\} \\ &= \beta E_{\frac{a^*}{\beta}}^{*}[v|s].\end{aligned}$$

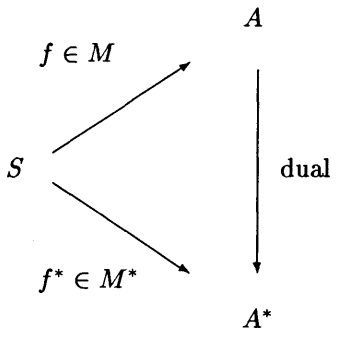
Let $\widetilde{B}(S)$ be the set of all real-valued measurable functions and let M^* be the set of all measurable mappings f^* from S to A^*, that is, $f^*(s) \in A^*$ for every $s \in S$. For each $s \in S$ and $f^* \in M^*$, we define a dual operator $T_{f^*}^{*}$ on $\widetilde{B}(S)$ by

$$T_{f^*}^{*} v(s) = -r^*(s, -f^*(s)) - \beta E_{\frac{f^*}{\beta}}^{*}[v|s], \tag{4.2}$$

and, further, define a dual operator T^* on $\widetilde{B}(S)$ by

$$T^* v(s) = \sup_{f^* \in M^*} T_{f^*}^{*} v(s). \tag{4.3}$$

Roughly, for each $s \in S$, we shall show the relation between the Banach space A and the dual space A^*, and the relation between the mapping f and the mapping f^* by the following diagram :



Here, we give the dual form of the dynamic model (2.1) as follows

$$(S, A^*, q, -r^*, \beta). \tag{4.4}$$

(i) S is the same as the original model (2.1), namely, the *state space of the dual model.*

(ii) A^* is the dual space of the action space A, namely, the *action space of the dual model.*

(iii) q is the *transition probability measure* given in (2.1), that is, the law of motion of the dual model is given by q.

(iv) r^* is the conjugate function of the loss function r in (2.1) and, for each $s \in S$, $-r^*(s,\cdot) : A^* \to \underline{\mathbb{R}}$ is called the *one-stage dual reward function*, where $\underline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$.

(v) β is the same discount factor as the original model (2.1).

A policy π^*, which we call a dual policy, for the dual model (4.4) is defined by infinite sequence $\pi^* = (f_0^*, f_1^*, \cdots, f_k^*, \cdots)$, each component f_k^* of which belongs to M^*. We denote that Π^* is the set of all dual policies. Then, under the discount factor β and the zero function $v_0 \equiv 0$, if a dual policy $\pi^* = (f_0^*, f_1^*, \cdots, f_k^*, \cdots)$ is used at an initial state s, then, using (4.1) and (4.2), we get the discounted reward up to the 2-stage as

$$\begin{aligned} I_{\pi^*}^{*(2)}(s) &= T_{f_0^*}^* T_{f_1^*}^* v_0(s) \\ &= -r^*(s, -f_0^*(s)) - \beta E_{\frac{f_0^*}{\beta}}^* [T_{f_1^*}^* v_0 | s]. \end{aligned}$$

By the successive method, the discounted reward up to the t-stage is written by

$$I_{\pi^*}^{*(t)}(s) = T_{f_0^*}^* T_{f_1^*}^* \cdots T_{f_t^*}^* v_0(s),$$

and, further the total discounted reward is given by

$$I_{\pi^*}^*(s) = \lim_{t \to \infty} I_{\pi^*}^{*(t)}(s).$$

Thus, a dual optimization problem for **(MP)** is given by

(DMP) maximize $I_{\pi^*}^*(s)$ subject to $\pi^* \in \Pi^*$,

For this problem at an initial state $s \in S$, we define an optimal dual value $I^*(s)$ as

$$I^*(s) = \sup_{\pi^* \in \Pi^*} I_{\pi^*}^*(s)$$

and, if $I^*(s) = I_{\overline{\pi}^*}^*(s)$, the dual policy $\overline{\pi}^* \in \Pi^*$ is said to be *optimal* for an initial state s. Then, we can get the optimal value for the dual model (4.4) as

$$I^*(s) = \lim_{t \to \infty} (T^*)^t v_0(s).$$

From now, we shall study the relations between the original model (2.1) and the dual model (4.4).

LEMMA 4.2. *For each $s \in S$ and $v \in B(S)$, $Tv(s) \geq T^*v(s)$.*

Proof. By virtue of Fenchel's inequality, for each $f \in M$ and $f^* \in M^*$, we have

$$\begin{aligned} T_f v(s) - T^*_{f^*} v(s) &= r(s, f(s)) + \beta E_f[v|s] + r^*(s, -f^*(s)) + \beta E^*_{\frac{f^*}{\beta}}[v|s] \\ &= r(s, f(s)) + r^*(s, -f^*(s)) + \beta E_f[v|s] + \beta E^*_{\frac{f^*}{\beta}}[v|s] \\ &\quad \text{(by Fenchel's inequality)} \\ &\geq \langle f(s), -f^*(s)\rangle + \beta \left\langle f(s), \frac{f^*(s)}{\beta} \right\rangle \\ &= -\langle f(s), f^*(s)\rangle + \langle f(s), f^*(s)\rangle \\ &= 0. \end{aligned}$$

This implies that

$$\inf_{f \in M} T_f v(s) - \sup_{f^* \in M^*} T^*_{f^*} v(s) \geq 0.$$

Thus, we get $Tv(s) \geq T^* v(s)$ for all $s \in S$. This lemma is proved. □

For each $s \in S$ and $v \in \widetilde{B}(S)$, we introduce the following notations for the domains of $r^*(s, \cdot)$ and $E^*_{\cdot/\beta}[v|s]$:

$$A^*_s = \left\{ a^* \in A^* \;\middle|\; r^*(s, a^*) < +\infty \right\} \subset A^*,$$
$$B^*_s(v) = \left\{ b^* \in A^* \;\middle|\; E^*_{\frac{b^*}{\beta}}[v|s] < +\infty \right\} \subset A^*.$$

ASSUMPTION 4.3. There exists an $f^* \in M^*$ such that, for all $s \in S$ and $v \in \widetilde{B}(S)$,

$$f^*(s) \in \operatorname{int}(B^*_s(v) \cap A^*_s),$$

where $\operatorname{int}(B^*_s(v) \cap A^*_s)$ is the set of all interior points of $B^*_s(v) \cap A^*_s$.

Under this assumption, there exists an $f^* \in M^*$ such that $T^*_{f^*} : \widetilde{B}(S) \to \widetilde{B}(S)$. Thus, we have $T^* v(s) > -\infty$. For each $s \in S$ and $v \in \widetilde{B}(S)$, we introduce a mapping $\psi_s : A^*_s \times B^*_s(v) \to \mathbb{R} \times A^*$ defined by

$$\psi_s(a^*, b^*) = \left(-r^*(s, a^*) - \beta E^*_{\frac{b^*}{\beta}}[v|s], a^* + b^* \right)$$

together with the set $F_s(v) \equiv \psi_s(A^*_s \times B^*_s(v)) - [0, \infty) \times \{\theta^*\}$ constructed by difference of vectors, where θ^* is the zero vector in A^*.

LEMMA 4.4. *Under assumption 4.1, suppose in addition that, for each $s \in S$ and $v \in B(S)$, the zero vector θ in the action space A belongs to $\operatorname{int} B_s(v)$. Then $F_s(v)$ is the convex and w^*-closed set in $\mathbb{R} \times A^*$, where $\mathbb{R} \times A^*$ is a dual space of $\mathbb{R} \times A$.*

Proof. Since $r^*(s, \cdot)$ and $E^*_{\cdot/\beta}[v|s]$ are convex functions, A^*_s and $B^*_s(v)$ are convex sets. Therefore, by the convexity of these conjugate functions, it is easily shown that $F_s(v)$ is the convex set.

In order to show that $F_s(v)$ is w*-closed, we consider a sequence of elements (v_s^n, r_s^n) in $F_s(v)$ converging (w*-topology) to (v_s^*, r_s^*) in $\mathbb{R}\times A^*$. Thus, from the definition of $F_s(v)$, there exist $p_s^n \in A_s^*$ and $q_s^n \in B_s^*(v)$ such that, for some $c \in [0, \infty)$,

$$v_s^n = -r^*(s, p_s^n) - \beta E^*_{\frac{q_s^n}{\beta}}[v|s] - c, \qquad r_s^n = p_s^n + q_s^n. \tag{4.5}$$

From (4.5), we have

$$v_s^n \leq -r^*(s, p_s^n) - \beta E^*_{\frac{q_s^n}{\beta}}[v|s], \qquad r_s^n = p_s^n + q_s^n. \tag{4.6}$$

Since $\theta \in \operatorname{int} B_s(v)$, there exists a ball $B_\varepsilon = \{q_s \in B \mid \|q_s\| < \varepsilon\}$ such that $B_\varepsilon \subset B_s(v) \subset A(s)$. Thus, for all $z \in A$, there exist two vectors $a, b \in B_s(v)$ such that $(\varepsilon/\|z\|)z = a - b$. Therefore, using Fenchel's inequality, we obtain

$$\begin{aligned}
\frac{\varepsilon}{\|z\|}\langle z\,,\, q_s^n\rangle &= \langle a - b\,,\, q_s^n\rangle \\
&= \langle a\,,\, q_s^n\rangle - \langle b\,,\, q_s^n\rangle \\
&\quad (\text{by } r_s^n = p_s^n + q_s^n \text{ in (4.6)}) \\
&= \langle a\,,\, r_s^n - p_s^n\rangle - \langle b\,,\, q_s^n\rangle \\
&= \langle a\,,\, r_s^n\rangle - \langle a\,,\, p_s^n\rangle - \langle b\,,\, q_s^n\rangle \\
&\quad (\text{by Fenchel's inequality}) \\
&\geq \langle a\,,\, r_s^n\rangle - \{r(s,a) + r^*(s, p_s^n)\} - \beta\left\{E_{\frac{b}{\beta}}[v|s] + E^*_{\frac{q_s^n}{\beta}}[v|s]\right\} \\
&\quad (\text{by the definition of } v_s^n) \\
&\geq \langle a\,,\, r_s^n\rangle + v_s^n - r(s,a) - \beta E_{\frac{b}{\beta}}[v|s].
\end{aligned}$$

Since the sequences $\{v_s^n\}$ and $\{\langle a\,,\, r_s^n\rangle\}$ converge to v_s^* and $\langle a\,,\, r_s^*\rangle$, respectively, we arrive at $\inf_{n\geq 1}\langle z\,,\, q_s^n\rangle > -\infty$ for all $z \in A$. According to uniform boundedness theorem, we get that $\{q_s^n\}_{n\geq 1}$ is bounded. Thus, from Alaoglu's theorem, it is w*-compact. See Chapter 5 in Luenberger (1969) for details. Hence, there exists a subsequence $\{q_s^{n(j)}\}_{j\geq 1}$ of $\{q_s^n\}_{n\geq 1}$ which converges to $q_s^* \in A^*$ in w*-topology, and consequently, a subsequence $p_s^{n(j)} = r_s^{n(j)} - q_s^*$ converges to $p_s^* = r_s^* - q_s^*$ in w*-topology.

From the construction of conjugate functions, $-r^*(s, \cdot)$ and $-E^*_{\cdot/\beta}[v|s]$ are w*-upper semicontinuous. Therefore, we get

$$\begin{aligned}
-r^*(s, p_s^*) - \beta E^*_{\frac{q_s^*}{\beta}}[v|s] &\geq \limsup_{j\to\infty}\left\{-r^*(s, p_s^{n(j)})\right\} + \limsup_{j\to\infty}\left\{-\beta E^*_{\frac{q_s^{n(j)}}{\beta}}[v|s]\right\} \\
&\geq \limsup_{j\to\infty}\left\{-r^*(s, p_s^{n(j)}) - E^*_{\frac{q_s^{n(j)}}{\beta}}[v|s]\right\} \geq \lim_{n\to\infty} v_s^n = v_s^*.
\end{aligned}$$

Thus, we obtain $-r^*(s, p_s^*) - \beta E^*_{q_s^*/\beta}[v|s] \geq v_s^*$ and $r_s^* = p_s^* + q_s^*$. It implies that (v_s^*, r_s^*) belongs to $F_s(v)$, which shows that $F_s(v)$ is w*-closed. □

In order to prove the main theorems, we introduce a set-valued function $G_s(v) : S \rightsquigarrow A^*$ defined by

$$G_s(v) = \{a^* \in A^* | \psi_s(-a^*, a^*) \in F_s(v)\}, \text{ for each } v \in \widetilde{B}(S).$$

From Lemma 4.4, $G_s(v)$ is the w*-closed set-valued function.

ASSUMPTION 4.5. For each $v \in \widetilde{B}(S)$, the set-valued function $G_s(v) : S \rightsquigarrow A^*$ is lower measurable, that is, for each open set O in A^*, the set $\{s \in S | G_s(v) \cap O \neq \phi\}$ belongs to $\mathcal{B}(S)$, where $\mathcal{B}(S)$ is the Borel field of S.

The following useful characterization of lower measurable w*-closed set-valued functions is given in Castaing and Valadier (1977), and Himmelberg (1975).

LEMMA 4.6. *For each $v \in \widetilde{B}(S)$, the w*-closed set-valued function $G_s(v) : S \rightsquigarrow A^*$ is lower measurable if and only if, for each $s \in S$, there exists a countable set $\{f_n^*\}_{n=1,2,\cdots} \subset M^*$ such that $G_s(v)$ is equal to the closure of $\{f_n^*(s)\}$ for every $s \in S$.*

THEOREM 4.7. *Under the assumptions 4.1, 4.3, and 4.5, suppose that, for each $s \in S$ and $v \in B(S)$, $\theta \in \operatorname{int} B_s(v)$ and*

$$(Tv(s), \theta^*) \in F_s(v). \tag{4.7}$$

Then, it holds that :

(i) $Tv(s) = T^*v(s)$;

(ii) *there exists an $f^* \in M^*$ such that*

$$-r^*(s, -f^*(s)) - \beta E^*_{\frac{f^*}{\beta}}[v|s] = T^*v(s)$$

for each $s \in S$ and $v \in \widetilde{B}(S)$.

Proof. Since we assume that, for each $s \in S$ and $v \in B(S)$, $(Tv(s), \theta^*) \in F_s(v)$ under the assumptions 4.1 and 4.3, it follows that there exists a $q_s^* \in A^*$ such that from (4.7)

$$\begin{aligned} Tv(s) &\leq -r^*(s, -q_s^*) - \beta E^*_{\frac{q_s^*}{\beta}}[v|s] \\ &\quad \text{(by the definition of } T^*) \\ &\leq T^*v(s) \\ &\quad \text{(by Lemma 4.2)} \\ &\leq Tv(s), \end{aligned}$$

for each $s \in S$ and $v \in B(S)$. Consequently, from Lemma 4.6, there exists an $f^* \in M^*$ satisfying that $Tv(s) = T^*v(s) = -r^*(s, -f^*) - \beta E^*_{f^*/\beta}[v|s]$, which completes the proof of the theorem. □

REMARK. If we have the following assumption, we can show that (4.7) in Theorem 4.7 holds.

ASSUMPTION 4.8. For each $s \in S$ and $v \in B(S)$:

(i) $r(s,\cdot)$ is convex and lower semicontinuous on $A(s)$;

(ii) $E_\cdot[v|s]$ is convex and lower semicontinuous on $B_s(v) \subset A(s)$.

We will sketch an outline of its proof as follows. Now we suppose $(Tv(s),\theta^*) \notin F_s(v)$. Since, from Lemma 4.4, $F_s(v)$ is convex and w*-closed, and $\mathbb{R}\times A^*$ is the dual space of $\mathbb{R}\times A$, $(Tv(s),\theta^*)$ may be strictly separated from $F_s(v)$ (see Chapter 5 in Luenberger (1969)) Thus, there exists $(\alpha,p_s) \in \mathbb{R}\times A$ and $\varepsilon > 0$ such that

$$\begin{aligned} \alpha Tv(s) &\geq \sup_{\substack{(p_s^*,q_s^*)\in A_s^*\times B_s^* \\ c\in\mathbb{R}_+}} \left[\left\langle (\alpha,p_s)\,,\,(-r^*(s,p_s^*) - \beta E^*_{\frac{q_s^*}{\beta}}[v|s] - c, p_s^* + q_s^*)\right\rangle\right] + \varepsilon \\ &= \sup_{(p_s^*,q_s^*)\in A_s^*\times B_s^*} \left[\alpha\{-r^*(s,p_s^*) - \beta E^*_{\frac{q_s^*}{\beta}}[v|s]\} + \langle p_s\,,\,p_s^* + q_s^*\rangle\right] \\ &\quad - \inf_{c\in\mathbb{R}_+} \alpha c + \varepsilon. \end{aligned} \tag{4.8}$$

Since $-\inf_{c\in\mathbb{R}_+}\alpha c$ is bounded from above in (4.8), it follows that $\inf_{c\in\mathbb{R}_+}\alpha c = 0$ and that α is positive or 0. Furthermore, it cannot be zero, because, if α is zero, we have

$$\sup_{(p_s^*,q_s^*)\in A_s^*\times B_s^*} \langle p_s\,,\,p_s^* + q_s^*\rangle + \varepsilon \leq 0.$$

Since there exist $p_s^* \in B_s^*(v)$ and $q_s^* \in B_s^*(v)$ satisfying $p_s^* + q_s^* = \theta^*$, we arrive at $0 \geq \varepsilon$, which is impossible. Thus, α is positive. On the other hand, since $r(s,\cdot)$ and $E_\cdot[v|s]$ are convex and lower semicontinuous, we obtain $r^{**}(s,\cdot) = r(s,\cdot)$ and $E_\cdot^{**}[v|s] = E_\cdot[v|s]$ for all $s \in S$ and $v \in B(S)$, where $r^{**}(s,\cdot)$ and $E_\cdot^{**}[v|s]$ are the biconjugates of $r(s,\cdot)$ and $E_\cdot[v|s]$, respectively. Dividing both sides of (4.8) by $\alpha > 0$ and putting $\eta = \varepsilon/\alpha$, we have

$$\begin{aligned} Tv(s) &\geq \sup_{(p_s^*,q_s^*)\in A_s^*\times B_s^*} \left[-r^*(s,p_s^*) + \langle p_s\,,\,p_s^*\rangle - \beta E^*_{\frac{q_s^*}{\beta}}[v|s] + \langle p_s\,,\,q_s^*\rangle\right] + \eta \\ &= \sup_{p_s^*\in A_s^*} [\langle p_s\,,\,p_s^*\rangle - r^*(s,p_s^*)] + \sup_{q_s^*\in B_s^*} \left[\langle p_s\,,\,q_s^*\rangle - \beta E^*_{\frac{q_s^*}{\beta}}[v|s]\right] + \eta \\ &= \sup_{p_s^*\in A^*} [\langle p_s\,,\,p_s^*\rangle - r^*(s,p_s^*)] + \beta \sup_{\frac{q_s^*}{\beta}\in A^*} \left[\left\langle p_s\,,\,\frac{q_s^*}{\beta}\right\rangle - E^*_{q_s^*}[v|s]\right] + \eta \\ &= r^{**}(s,p_s) + \beta E^{**}_{p_s}[v|s] + \eta \\ &= r(s,p_s) + \beta E_{p_s}[v|s] + \eta. \end{aligned} \tag{4.9}$$

Thus, there exists an $f \in M$ such that (4.9) is written as

$$Tv(s) \geq T_f v(s) + \eta. \tag{4.10}$$

Since $Tv(s) \leq T_f v(s)$ for all $f \in M$, thus (4.10) is impossible. Therefore, it follows that (4.7) in Theorem 4.7 holds.

THEOREM 4.9. *Suppose that the same assumptions as in Theorem 4.7 hold. For $I(s)$ of the original model and the optimal dual value $I^*(s)$ of the dual one, we have for each s*

$$I(s) = I^*(s)$$

and, moreover, there exists an optimal dual policy $\overline{\pi}^* \in \Pi^*$, that is,

$$I^*_{\overline{\pi}^*}(s) = I^*(s). \tag{4.11}$$

Proof. From Theorem 4.7, it follows that, for each $s \in S$ and $v \in B(S)$, there exists an $f^* \in M^*$ such that $T^*v(s) = T^*_{f^*}v(s) = Tv(s)$. Consequently, we get that for each t there exists a finite sequence $\{f^*_k\}_{k=0,\cdots,t} \subset M^*$ such that

$$\begin{aligned}(T^*)^{t+1}v_0(s) &= T^*_{f^*_0}\cdots T^*_{f^*_t}v_0(s) \\ &= (T)^{t+1}v_0(s).\end{aligned} \tag{4.12}$$

Thus, from (4.12), there exists a $\overline{\pi}^* = (f^*_0, f^*_1, \cdots)$ such that

$$I^*(s) = \lim_{t\to\infty}(T^*)^t v_0(s) = T^*_{\overline{\pi}^*}v_0(s) = \lim_{t\to\infty}(T)^t v_0(s) = I(s).$$

This dual policy $\overline{\pi}^*$ is optimal for **(DMP)** and, at the same time, the dual optimal value $I^*_{\overline{\pi}^*}(s)$ is equal to $I(s)$ for **(MP)**. □

COROLLARY 4.10. *In addition to the condition in Theorem 4.9, suppose that the assumption 4.8 holds, and that for each $s \in S$ and $v \in B(S)$, $\theta \in \operatorname{int} B_s(v)$. Then, there exist $\overline{\pi} \in \Pi$ and $\overline{\pi}^* \in \Pi^*$ such that for all $s \in S$, it follows that*

$$I^*(s) = I^*_{\overline{\pi}^*}(s) = I_{\overline{\pi}}(s) = I(s). \tag{4.13}$$

Proof. Under assumption *4.8*, we have that for each $s \in S$ and $v \in B(S)$,

$$r^{**}(s,a) = r(s,a), \qquad \text{and} \qquad E^{**}_{\cdot}[v|s] = E_{\cdot}[v|s].$$

By a similar argument as in Theorem 4.9, it follows that there exist $\overline{\pi} \in \Pi$ and $\overline{\pi}^* \in \Pi^*$ such that for all $s \in S$, satisfy (4.13). □

Acknowledgement

The authors would like to thank the referee for his very helpful comments and suggestions.

References

Aubin,J.P.(1982), Mathematical Methods of Game and Economic Theory, Revised Edition, North-Holland, Amsterdam.

Aubin,J.P.(1993), Optima and Equilibria, Springer-Verlag, New York.

Badler,E.J.(1989), On compactness of the space of policies in stochastic dynamic programming, Stochastic Proc. Appl. 32, 141-150.
Bertsekas,D.P. and Shreve,S.E.(1978), Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York.
Blackwell,D.(1965), Discounted dynamic programming, Ann. Math. Statist. 36, 226-235.
Borkar,V.S.(1991), Topics in Controlled Markov Chains, Longman Scientific & Technical, Longman Group UK limited.
Castaing,C. and Valadier,M.(1977), Convex Analysis and Measurable Multifunctions, Lecture Notes in Mathematics 580, Springer-Verlag, Berlin, 1977.
Dynkin,E.B. and Yushkevich,A.A.(1979), Controlled Markov Processes, Springer-Verlag, Berlin.
Iwamoto,S.(1983), Reverese function, reverse program, and reverse theorem in mathematical programming, J. Math. Anal. Appl. 95, 1-19.
Iwamoto,S.(1984), A dynamic inversion of the classical variational problems, J. Math. Anal. Appl. 95, 354-374.
Himmelberg,C.J.(1975), Measurable relations, Fund. Math. 87, 53-72.
Luenberger,D.G.(1969), Optimization by Vector Space Methods, John Wiley & Sons, inc.
Rockafellar,R.T.(1966), Extension of Fenchel's duality theorem for convex functions. Duke Math. J. 33, 81-89.
Schäl,M.(1975), On dynamic programming: Compactness of the space of policies, Stochastic Processes Appl. 3, 345-364.
Strauch,R.(1966), Negative dynamic programming, Ann. Math. Statist., 37, 871-890.
Tanaka,K.(1991), On discounted dynamic programming with constraints, J. Math. Anal. Appl. 155, 264-277.
Tanaka,K., Hoshino,M. and Kuroiwa,D.(1995), On an $\varepsilon-$optimal policy of discrete time stochastic control processes, Bulletin of Informatics and Cybernetics, 27, 107-119.

Received November 12, 1996
Revised January 31, 1997