Fuzzy Optimality Equations for Perceptive MDPs

Masami Kurano[1] Masami Yasuda[2] Jun-ichi Nakagami[2] Yuji Yoshida[3]

1.Faculty of Education, Chiba University, Chiba 263-8522 Japan

2.Faculty of Science, Chiba University, Chiba 263-8522 Japan

3.The University of Kitakyushu, Kitakyushu 802-8577 Japan

Email: kurano@math.e.chiba-u.ac.jp, yasuda@math.s.chiba-u.ac.jp

nakagami@math.s.chiba-u.ac.jp, yoshida@kitakyu-u.ac.jp

ABSTRACT: This paper is a sequel to Kurano et al [9], [10], in which the fuzzy perceptive models for optimal stopping or discounted Markov decision process are proposed and the methods of computing the corresponding fuzzy perceptive values are given.

Here, we deal with the average case for Markov decisin processes with fuzzy perceptive transition matrices and characterize the optimal average expected reward, called the average perceptive value, by a fuzzy optimality equation. Also, we give a numerical example.

Keywords : Fuzzy perceptive model, Markov decision process, average criterion, fuzzy perceptive value, optimal policy function.

1. Introduction and notation

In a real application of such a mathematical model as a Markov decision process (MDP), it often occurs that the required data is linguistically or roughly perceived (for example, the probability of the transition from one state to another is about 0.3 or considerably larger than 0.8, etc.).

A possible way of handling such a perception-based information is to use the fuzzy set (cf. [4], [17]), whose membership function describes the level of the perception of the required data. If the fuzzy perception of the transition matrices in MDPs is given, how can we estimate the future expected reward, called a fuzzy perceptive value, in advance of our actual decision, under the condition that we can know the true value of the transition matrices immediately before our decision making. The concept of fuzzy perceptive values is the same as the perceptive value (possibility distribution) of the objective function under the possibility constraints proposed by Zadeh [18] using a generalized extension principle.

In our previous works [9], [10], we have given the perceptive models for an optimal stopping or discounted MDPs and the corresponding fuzzy perceptive values are characterized and calculated by the corresponding fuzzy optimality equations. As for MDPs, the average case was not treated there.

The objective of this paper is to formulate the perceptive model for average reward MDPs and derive the average fuzzy optimality equation by which the average fuzzy perceptive values are obtained. In order to guarantee the ergodicity of the process, we impose the minorization condition (cf. [12]). Also, as a numerical example, a machine maintenance problem is considered.

In remainder of this section, we will give some notation and fundamental results on average reward MDPs, from which the fuzzy perceptive model is formulated in the sequel. For non-perception approaches to MDPs with fuzzy imprecision refer to [8].

Let $\mathbb{R}, \mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ be the sets of real numbers, real n-dimensional vectors and real $m \times n$ matrices, respectively. The sets $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ are endowed with the norm $\|\cdot\|$, where for $x = (x(1), x(2), \dots, x(n)) \in \mathbb{R}^n$, $\|x\| = \sum_{j=1}^n |x(j)|$ and for $y = (y_{ij}) \in \mathbb{R}^{m\times n}$, $\|y\| = \max_{1\le i\le m} \sum_{j=1}^n |y_{ij}|$.

For any set X, let $\mathcal{F}(X)$ be the set of all fuzzy sets $\widetilde{x} :\to [0,1]$. The α-cut of $\widetilde{x} \in \mathcal{F}(X)$ is given by $\widetilde{x}_\alpha := \{x \in X \mid \widetilde{x}(x) \ge \alpha\}$ $(\alpha \in (0,1])$ and $\widetilde{x}_0 := \mathrm{cl}\{x \in X \mid \widetilde{x}(x) > 0\}$, where cl is a closure of a set. Let $\widetilde{\mathbb{R}}$ be the set of all fuzzy numbers, i.e., $\widetilde{r} \in \widetilde{\mathbb{R}}$ means that $\widetilde{r} \in \mathcal{F}(\mathbb{R})$ is normal, upper semicontinuous and fuzzy convex and has a compact support. Let $\mathbb{C}$ be the set of all bounded and closed intervals of $\mathbb{R}$. Then, for $\widetilde{r} \in \mathcal{F}(\mathbb{R})$, it holds that $\widetilde{r} \in \widetilde{\mathbb{R}}$ if and only if $\widetilde{r}$ normal and $\widetilde{r}_\alpha \in \mathbb{C}$ for $\alpha \in [0,1]$). So, for $\widetilde{r} \in \widetilde{\mathbb{R}}$, we write $\widetilde{r}_\alpha = [\widetilde{r}_\alpha^-, \widetilde{r}_\alpha^+]$ $(\alpha \in [0,1])$.

The binary relation $\preccurlyeq$ on $\mathcal{F}(\mathbb{R})$ is defined as follows: For $\widetilde{r}, \widetilde{s} \in \mathcal{F}(\mathbb{R})$, $\widetilde{r} \preccurlyeq \widetilde{s}$ if and only if (i) for any $x \in \mathbb{R}$, there exists $y \in \mathbb{R}$ such that $x \le y$ and $\widetilde{r}(x) \le \widetilde{s}(y)$; (ii) for any $y \in \mathbb{R}$, there exists $x \in \mathbb{R}$ such that $x \le y$ and $\widetilde{s}(y) \le \widetilde{r}(x)$: Obviously, the binary relation $\preccurlyeq$ satisfies the axioms of a partial order relation on $\mathcal{F}(\mathbb{R})$ (cf. [7], [16]).

For $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$, $\widetilde{\max}\{\widetilde{r}, \widetilde{s}\}$ and $\widetilde{\min}\{\widetilde{r}, \widetilde{s}\}$ are defined by

$$\widetilde{\max}\{\widetilde{r}, \widetilde{s}\}(y) := \sup_{\substack{x_1, x_2 \in \mathbb{R} \\ y = x_1 \vee x_2}} \{\widetilde{r}(x_1) \wedge \widetilde{s}(x_2)\} \quad (y \in \mathbb{R}),$$

and

$$\widetilde{\min}\{\widetilde{r}, \widetilde{s}\}(y) := \sup_{\substack{x_1, x_2 \in \mathbb{R} \\ y = x_1 \wedge x_2}} \{\widetilde{r}(x_1) \wedge \widetilde{s}(x_2)\} \quad (y \in \mathbb{R}),$$

where $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for any $a, b \in \mathbb{R}$. It is easy proved that for $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$, $\widetilde{\max}\{\widetilde{r}, \widetilde{s}\} \in \widetilde{\mathbb{R}}$ and $\widetilde{\min}\{\widetilde{r}, \widetilde{s}\} \in \widetilde{\mathbb{R}}$.

Also, for $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$, the following (i)–(iv) are equivalent (cf. [7]): (i) $\widetilde{r} \preccurlyeq \widetilde{s}$; (ii) $\widetilde{r}_\alpha^- \le \widetilde{s}_\alpha^-$ and $\widetilde{r}_\alpha^+ \le \widetilde{s}_\alpha^+$ $(\alpha \in [0,1])$; (iii) $\widetilde{\max}\{\widetilde{r}, \widetilde{s}\} = \widetilde{s}$; (iv) $\widetilde{\min}\{\widetilde{r}, \widetilde{s}\} = \widetilde{r}$.

For any $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$,

$$(\widetilde{r} + \widetilde{s})(y) := \sup_{\substack{x_1, x_2 \in \mathbb{R} \\ y = x_1 + x_2}} \{\widetilde{r}(x_1) \wedge \widetilde{s}(x_2)\} \quad (y \in \mathbb{R}),$$

When $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$, it holds (cf. [4]) that $\widetilde{r} + \widetilde{s} \in \widetilde{\mathbb{R}}$ and $(\widetilde{r} + \widetilde{s})^-_\alpha = \widetilde{r}^-_\alpha + \widetilde{s}^-_\alpha$ and $(\widetilde{r} + \widetilde{s})^+_\alpha = \widetilde{r}^+_\alpha + \widetilde{s}^+_\alpha$ ($\alpha \in [0, 1]$).

We denote by $\mathbb{R}_+$ and $\mathbb{R}^n_+$ the subsets of entrywise non-negative elements in $\mathbb{R}$ and $\mathbb{R}^n$ respectively. Let $\mathbb{C}_+$ be the set of all bounded and closed intervals of $\mathbb{R}_+$ and $\mathbb{C}^n_+$ the set of all n-dimensional vectors whose elements are in $\mathbb{C}_+$.

We have the following.

Lemma 1.1 ([6]) *For any non-empty convex and compact set $G \subset \mathbb{R}^n_+$ and $D = (D_1, D_2, \ldots, D_n) \in \mathbb{C}^n_+$, it holds that*

$$GD = \{g \cdot d \mid g \in G, d \in D\} \in \mathbb{C}_+$$

where $g \cdot d = \sum_{j=1}^n g_j d_j$ for $g = (g_1, g_2, \ldots, g_n) \in \mathbb{R}^n_+$ and $d = (d_1, d_2, \ldots, d_n) \in D$.

Here, we define average reward MDPs whose extension to the fuzzy perceptive model will be done in Section 2. Consider finite state and action spaces, S and A, containing $n < \infty$ and $k < \infty$ elements with $S = \{1, 2, \ldots, n\}$ and $A = \{1, 2, \ldots, k\}$.

Let $\mathcal{P}(S) \subset \mathbb{R}^n$ and $\mathcal{P}(S|SA) \subset \mathbb{R}^{n \times nk}$ be the sets of all probabilities on S and conditional probabilities on S given $S \times A$, that is,

$$\begin{aligned} \mathcal{P}(S) &:= \{q = (q(1), q(2), \ldots, q(n)) \mid \\ &\qquad q(i) \geq 0, \sum_{i=1}^n q(i) = 1, \ i \in S\}, \\ \mathcal{P}(S|SA) &:= \{Q = (q_{ia}(\cdot) : i \in S, a \in A) \mid \\ &\qquad q_{ia}(\cdot) = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n)) \\ &\qquad \in \mathcal{P}(S), \ i \in S, a \in A\}. \end{aligned}$$

For any $Q = (q_{ia}(\cdot)) \in \mathcal{P}(S|SA)$, we define a controlled dynamic system $\mathcal{M}(Q)$, called a Markov decision process(MDP), specified by $\{S, A, Q, r\}$, where $r : S \times A \to \mathbb{R}_+$ is an immediate reward function.

When the system is in state $i \in S$ and action $a \in A$ is taken, then the system moves to a new state $j \in S$ selected according to $q_{ia}(\cdot)$ and the reward $r(i, a)$ is obtained. The process is repeated from the new state $j \in S$.

The sample space is the product space $\Omega = (S \times A)^\infty$ such that the projections X_t, Δ_t on the factors S, A describe that the state and the action at the t-th time of the process ($t \geq 0$).

A policy $\pi = (\pi_1, \pi_2, \ldots)$ is a sequence of conditional probabilities $\pi_t(A|x_0, a_0, \ldots, x_t) = 1$ for all histories $(x_0, a_0, \ldots, x_t) \in (S \times A)^t \times S$. The set of all policies is denoted by Π. A policy $\pi = (\pi_0, \pi_1, \ldots)$ is called randomized stationary if there exists a conditional probability $\gamma = (\gamma(\cdot|i), i \in S)$ given S for which $\pi(\cdot|x_0, a_0, \ldots, x_t) = \gamma(\cdot|x_t)$ for all $t \geq 0$ and $(x_0, a_0, \ldots, x_t) \in (S \times A)^t \times S$. Such a policy is simply denoted by γ. We denote by F the set of functions from S to A. A randomized stationary policy γ is called stationary if there exists a function $f \in F$ such that $\gamma(\{f(i)\}|i) = 1$, for all $i \in S$, which is denoted by f.

For each $\pi \in \Pi$, stating state $X_0 = i$ and transition matrix $Q \in \mathcal{P}(S|SA)$, the probability measure $P_\pi(\cdot|X_0 = i, Q)$ on Ω is defined in an usual way. The problem we are concerned with is the maximization of the long-run expected average reward per unit time, $\varphi(i, \pi|Q)$, which is defined, as a function of $Q \in \mathcal{P}(S|SA)$, by

$$\varphi(i, \pi|Q) = \liminf_{T \to \infty} \frac{1}{T} E_\pi(\varphi_T | X_0 = i, Q) \tag{1.1}$$

($i \in S, \pi \in \Pi$), where $E_\pi(\cdot|X_0 = i, Q)$ is the expectation w. r. t. $P_\pi(\cdot|X_0 = i, Q)$ and $\varphi_T = \sum_{t=0}^{T-1} r(X_t, \Delta_t)$ ($T \geq 1$).

For any $Q \in \mathcal{P}(S|SA)$, a policy π^* satisfying that

$$\varphi(i, \pi^*|Q) = \sup_{\pi \in \Pi} \varphi(i, \pi|Q) := \varphi(i|Q) \quad \text{for all } \ i \in S$$

is called to be Q-average optimal (simply Q-optimal).

In order to insure the ergodicity of the process, we introduce the minorization condition M (cf. [12]). We say that the transition matrix $Q = (q_{ia}(\cdot) : i \in S, a \in A) \in \mathcal{P}(S|SA)$ satisfies Condition M if

$$\delta(Q) := \min_{i, j \in S, \ a \in A} q_{ia}(j) > 0.$$

Let $\mathcal{B}(S)$ be the set of all functions $u := S \to \mathbb{R}$. For any $Q = (q_{ia}(\cdot) : i \in S, a \in A) \in \mathcal{P}(S|SA)$, we define the map $U\{Q\} : \mathcal{B}(S) \to \mathcal{B}(S)$ by

$$U\{Q\}u(i) = \max_{a \in A}\{r(i, a) + \sum_{j \in S}(q_{ia}(j) - \delta(Q))u(j) \tag{1.2}$$

for all $i \in S$. Then, if $Q \in \mathcal{P}(S|SA)$ satisfies Condition M, the map $U\{Q\}$ is a contraction, so that there exists a unique fixed point $v = v(Q) \in \mathcal{B}(S)$ such that

$$U\{Q\}v = v. \tag{1.3}$$

Putting $\varphi(Q) = \delta(Q) \sum_{j \in S} v(j)$ in (1.3), we obtain the optimality equation for the average expected reward:

$$v(Q)(i) + \varphi(Q) = \max_{a \in A}\{r(i, a) + \sum_{j \in S} q_{ia}(j) v(Q)(j)\}. \tag{1.4}$$

The following lemma follows from (2.3) (cf. [1, 3, 5, 13]).

Lemma 1.2 *Suppose that $Q \in \mathcal{P}(S|SA)$ satisfies Condition M. Then, $\varphi(Q) = \varphi(i|Q)$ (not depend on $i \in S$) and $f(i) \in A^*(i|Q)$ for all $i \in S$, f is Q-optimal, where $A^*(i|Q) := \{a \in A \mid$ a maximizes the right-hand side of (1.4)$\}$.*

Let $\mathcal{P}_M$ be the set of all $Q \in \mathcal{P}(S|SA)$ which satisfies Condition M. Then, we have the following used in the sequel.

Lemma 1.3 (cf. [14, 15]) *The optimal average reward $\varphi(Q)$ is continuous in $\mathcal{P}_M$.*

In Section 2, we define a fuzzy perceptive model for average reward MDPs, which is analyzed in Section 3 with a numerical example.

2. Fuzzy perceptive model

We define a fuzzy-perceptive model, in which fuzzy perception of the transition probabilities in MDPs is accommodated. In a concrete form, we use the fuzzy set on $\mathcal{P}(S|SA)$ whose membership function $\widetilde{Q}$ describes the perception value of the transition probability.

Firstly, for each $i \in S$ and $a \in A$, we give a fuzzy perception of $q_{ia}(\cdot) = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n))$, $\widetilde{Q}_{ia}(\cdot)$, which is a fuzzy set on $\mathcal{P}(S)$ and will be assumed to satisfy the following conditions (i)–(ii):

(i) (Normality) There exists a $q = q_{ia}(\cdot) \in \mathcal{P}(S)$ with $\widetilde{Q}_{ia}(q) = 1$;

(ii) (Convexity and compactness) The α-cut $\widetilde{Q}_{ia,\alpha}(\cdot) = \{q = q_{ia}(\cdot) \in \mathcal{P}(S) \mid \widetilde{Q}_{ia}(q) \geq \alpha\}$ is a convex and compact subset in $\mathcal{P}(S)$ ($\alpha \in [0, 1]$).

Secondly, from a family of fuzzy-perceptions $\{\widetilde{Q}_{ia}(\cdot) : i \in S, a \in A\}$, we define the fuzzy set $\widetilde{Q}$ on $\mathcal{P}(S|SA)$, called fuzzy perception of the transition probability in MDPs, as follows:

$$\widetilde{Q}(Q) = \min_{i \in S, a \in A} \widetilde{Q}_{ia}(q_{ia}(\cdot)), \tag{2.1}$$

where $Q = (q_{ia}(\cdot) : i \in S, a \in A) \in \mathcal{P}(S|SA)$.

The α-cut of the fuzzy perception $\widetilde{Q}$ is described explicitly in the following:

$$\begin{aligned} \widetilde{Q}_\alpha = \{Q = (q_{ia}(\cdot) : i \in S, a \in A) \in \mathcal{P}(S|SA) \mid \\ q_{ia}(\cdot) \in \widetilde{Q}_{ia,\alpha} \text{ for } i \in S, a \in A\} \\ = \prod_{i \in S, a \in A} \widetilde{Q}_{ia,\alpha} \quad (\alpha \in [0, 1]). \end{aligned} \tag{2.2}$$

Remark *For each $i \in S$ and $a \in A$, in place of giving the fuzzy perception $\widetilde{Q}_{ia}$ on $\mathcal{P}(S)$, it may be convenient to give the fuzzy set $\widetilde{q}_{ia}(j) \in \widetilde{\mathbb{R}}$ ($j \in S$) on $[0, 1]$, which represents the fuzzy perception of $q_{ia}(j)$ (the probability that the state moves to $j \in S$ when the action $a \in A$ is taken in state $i \in S$).*

Then, $\widetilde{Q}_{ia}(\cdot)$ is defined by

$$\widetilde{Q}_{ia}(q) = \min_{j \in S} \widetilde{q}_{ia}(j)(q_{ia}(j)), \tag{2.3}$$

where $q = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n)) \in \mathcal{P}(S)$.

For any fuzzy perception $\widetilde{Q}$ on $\mathcal{P}(S|SA)$, our fuzzy-perceptive model is denoted by $\mathcal{M}(\widetilde{Q})$, in which for any $Q \in \mathcal{P}(S|SA)$ the corresponding MDPs $\mathcal{M}(Q)$ is perceived with perception level $\widetilde{Q}(Q)$.

The map δ on $\mathcal{P}(S|SA)$ with $\delta(Q) \in \Pi$ for all $Q \in \mathcal{P}(S|SA)$ is called a policy function. The set of all policy functions will be denoted by Δ. For any $\delta \in \Delta$, the fuzzy perceptive reward $\widetilde{\varphi}$ is a fuzzy set on $\mathbb{R}$ denoted by

$$\widetilde{\varphi}(i, \delta)(x) = \sup_{\substack{Q \in \mathcal{P}(S|PS) \\ x = \varphi(i, \delta(Q)|Q)}} \widetilde{Q}(Q) \quad (i \in S). \tag{2.4}$$

The policy function $\delta^* \in \Delta$ is said to be optimal if $\widetilde{\varphi}(i, \delta) \preccurlyeq \widetilde{\varphi}(i, \delta^*)$ for all $i \in S$ and $\delta \in \Delta$, where the partial order $\preccurlyeq$ is defined in Section 1. If there exists an optimal policy function δ^*, we put $\widetilde{\varphi} = (\widetilde{\varphi}(1), \widetilde{\varphi}(2), \ldots, \widetilde{\varphi}(n))$ will be called a fuzzy perceptive value, where $\widetilde{\varphi}(i) = \widetilde{\varphi}(i, \delta^*)$ ($i \in S$).

Here, we can specify the fuzzy perceptive problem investigated in the next section: The problem is to find the optimal policy function δ^* and to characterize the fuzzy perceptive value.

3. Perceptive analysis

In this section, we derive a new fuzzy optimality relation for solving our perceptive problem. The sufficient condition for the fuzzy perceptive reward $\widetilde{\varphi}(i, \delta)$ to be a fuzzy number is given in the following lemma.

Lemma 3.1 *For any $\delta \in \Delta$, if $\varphi(i, \delta|Q)$ is continuous in $Q \in \widetilde{Q}_0$, then $\widetilde{\varphi}(i, \delta) \in \widetilde{\mathbb{R}}$, where $\widetilde{Q}_0$ is the 0-cut of $\widetilde{Q}$.*

Proof. ¿From normality of $\widetilde{Q}$, there exists $Q^* \in \mathcal{P}(S|SA)$ with $\widetilde{Q}(Q^*) = 1$, such that $\widetilde{\varphi}(i, \delta)(x^*) = 1$ for $x^* = \varphi(i, \delta|Q^*)$. For any $\alpha \in [0, 1]$, we observed that

$$\widetilde{\varphi}(i, \delta)_\alpha = \{\varphi(i, \delta|Q) \mid Q \in \widetilde{Q}_\alpha\}.$$

Since $\widetilde{Q}_\alpha$ is convex and compact, the continuity of $\varphi(i, \delta|\cdot)$ means the convexity and compactness of $\widetilde{\varphi}(i, \delta)_\alpha$ ($\alpha \in [0.1]$). □

Lemma 1.2 in Section 1 guarantees that for each $Q \in \mathcal{P}(S|SA)$ satisfying Condition M there exists a Q-optimal stationary policy f_* ($f_* \in F$). Thus, for each $Q \in \mathcal{P}(S|SA)$, we denote by $\delta^*(Q)$ the corresponding Q-optimal stationary policy, which is thought as a policy function.

Here we introduce a Minorization condition for the perceptive model $\mathcal{M}(\widetilde{Q})$. we say that $\widetilde{Q}$ on $\mathcal{P}(S|SA)$ satisfies Condition M if $\widetilde{Q}_0 \subset \mathcal{P}_M$, where $\widetilde{Q}_0$ is the 0-cut of $\widetilde{Q}$.

Lemma 3.2 *Suppose that $\widetilde{Q}$ satisfies Condition M. Then, $\varphi(i, \delta^*)$ is independent of $i \in S$ and $\widetilde{\varphi} := \widetilde{\varphi}(i, \delta^*) \in \widetilde{\mathbb{R}}$.*

Proof. By Lemma 1.2, $\widetilde{\varphi}(i, \delta^*|Q)$ is continuous in $\widetilde{Q}_0$, so that $\widetilde{\varphi}(i, \delta^*) \in \widetilde{\mathbb{R}}$ follows from Lemma 3.1. Also, from Lemma 1.1, $\varphi(i, \delta^*)$ is clearly independent of $i \in S$ □

Theorem 3.1 *The policy function δ^* is optimal.*

Proof. Let $\delta \in \Delta$. Since $\delta^*(Q)$ is Q-optimal, for any $Q \in \mathcal{P}(S|SA)$ it holds that

$$\varphi(i, \delta|Q) \leq \varphi(i, \delta^*|Q) \quad (i \in S). \tag{3.1}$$

For any $x \in \mathbb{R}$, let $\alpha := \widetilde{\varphi}(i, \delta)(x)$. Then, from the definition there exists $Q \in \widetilde{Q}_\alpha$ with $x = \varphi(i, \delta|Q)$. By (3,1), $y := \varphi(i, \delta^*|Q) \geq x$, which implies $\widetilde{\varphi}(i, \delta^*)(y) \geq \alpha$.

On the other hand, for $y \in \mathbb{R}$, let $\alpha := \widetilde{\varphi}(i,\delta^*)(y)$. Then, there exists $Q \in \widetilde{Q}_\alpha$ such that $y = \varphi(i,\delta^*|Q)$. ¿From (3.1), we have that $y \geq x := \varphi(i,\delta|Q)$. This implies $\widetilde{\varphi}(i,\delta|Q) \leq \alpha$. The above discussion yields that $\widetilde{\varphi}(i,\delta) \preccurlyeq \widetilde{\varphi}(i,\delta^*)$. □

¿From Lemma 3.2, $\widetilde{\varphi} \in \widetilde{\mathbb{R}}$ $(i \in S)$, so that we denote by $\widetilde{\varphi}_\alpha := [\widetilde{\varphi}_\alpha^-, \widetilde{\varphi}_\alpha^+(i)] \in \mathbb{C}$, the α-cut of $\widetilde{\varphi}$. In the following theorem, the fuzzy perceptive value $\widetilde{\varphi}$ is characterized by the fuzzy optimality relation.

Theorem 3.2 *Suppose that $\widetilde{Q} \in \mathcal{P}(S|SA)$ satisfies Condition M. Then, the fuzzy perceptive value $\widetilde{\varphi} \in \widetilde{\mathbb{R}}$ is a unique solution to the following fuzzy optimality relations:*

(3.2) $$\widetilde{v}_i + \widetilde{\varphi} = \widetilde{\max_{a\in A}}\{1_{\{r(i,a)\}} + \widetilde{Q}_{ia}\cdot\widetilde{v}\},$$

where $\widetilde{v} = (\widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{v}_n) \in \widetilde{\mathbb{R}}^n$ and $\widetilde{Q}_{ia}\cdot\widetilde{v}(x) = \sup \widetilde{Q}_{ia}(q) \wedge \widetilde{v}(\varphi)$ and the supremum is taken on the range $\{(q,\varphi) \mid x = \sum_{j=1}^n q(j)\varphi(j), q = (q(1), q(2), \ldots, q(n)) \in \mathcal{P}(S), \varphi = (\varphi(1), \varphi(2), \ldots, \varphi(n)) \in \mathbb{R}^n)\}$ and $\widetilde{v}(\varphi) = \widetilde{v}_1(\varphi(1)) \wedge \cdots \wedge \widetilde{v}_n(\varphi(n))$.

The α-cut expression of (3.2) is as follows:

(3.3)
$$\widetilde{v}_{i,\alpha}^- + \widetilde{\varphi}^- = \max_{a\in A}\{r(i,a) + \min_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} q_{ia}\cdot\widetilde{v}_\alpha^-\} \ (\alpha \in [0,1]);$$

(3.4)
$$\widetilde{v}_{i,\alpha}^+ + \widetilde{\varphi}^+ = \max_{a\in A}\{r(i,a) + \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} q_{ia}\cdot\widetilde{v}_\alpha^+\} \ (\alpha \in [0,1]);$$

where $\widetilde{v}_{i,\alpha} = [\widetilde{v}_{i,\alpha}^-, \widetilde{v}_{i,\alpha}^+]$, $\widetilde{\varphi}_\alpha^\mp = (\widetilde{\varphi}_{1,\alpha}^\mp, \ldots, \widetilde{\varphi}_{n,\alpha}^\mp)$, $\widetilde{v}_\alpha^\mp = (\widetilde{v}_{1,\alpha}^\mp, \ldots, \widetilde{v}_{n,\alpha}^\mp)$, and $q_{ia}\cdot\widetilde{v}_\alpha^\mp = \sum_{j\in S} q_{ia}(j)\widetilde{v}_{j,\alpha}^\mp$.

We note that α-cut of $\widetilde{Q}_{ia}\cdot\widetilde{v}$ in (3.2) is in $\mathbb{C}$ from Lemma 1.1, so that $\widetilde{Q}_{ia}\cdot\widetilde{v} \in \widetilde{\mathbb{R}}$. Thus, the right hand of (3.2) is well-defined.

Proof. Under Condition M, $\widetilde{Q}_0 \subset \mathcal{P}_M$, such that $\delta := \min_{Q\in\widetilde{Q}_0} \delta(Q) > 0$ and $q_{ia}(j) \geq \delta$ for all $q = (q_{ia}(\cdot)) \in \widetilde{Q}_{ia,\alpha}$ $(\alpha \in [0,1])$.

For any $\alpha \in [0,1]$, we define maps $\underline{U}^\alpha, \overline{U}^\alpha : \mathcal{B}(S) \to \mathcal{B}(S)$ by

(3.5)
$$\underline{U}^\alpha u(i) = \min_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \max_{a\in A}\{r(i,a) + \sum_{j=1}^n (q_{ia}(j)-\delta)u(j)\} \ (i \in S)$$

(3.6)
$$\overline{U}^\alpha u(i) = \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \max_{a\in A}\{r(i,a) + \sum_{j=1}^n (q_{ia}(j)-\delta)u(j)\} \ (i \in S),$$

for any $u \in \mathcal{B}(S)$.

Then, it is easily proved that the maps $\underline{U}^\alpha$ and $\overline{U}^\alpha$ are contractions with modulas $\beta = 1-\delta$ (<1). Thus, the unique fixed points exist for $\underline{U}^\alpha$ and $\overline{U}^\alpha$. Let denote the fixed points of $\underline{U}^\alpha$ and $\overline{U}^\alpha$ respectively by $\underline{v}_\alpha$ and $\overline{v}_\alpha \in \mathcal{B}(S)$. Also, by the same discussion as Lemma 4.2 in [10], we observe that $\underline{v}_\alpha$ and $\overline{v}_\alpha$ satisfy that

(3.7)
$$\underline{v}^\alpha(i) = \max_{a\in A}\{r(i,a) + \min_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^n (q_{ia}(j)-\delta)\underline{v}_\alpha(j)\} \ (i \in S),$$

(3.8)
$$\overline{v}^\alpha(i) = \max_{a\in A}\{r(i,a) + \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^n (q_{ia}(j)-\delta)\overline{v}_\alpha(j)\} \ (i \in S).$$

Putting $\varphi_\alpha^- = \sum_{j\in S} \underline{v}_\alpha(j)$ and $\varphi_\alpha^+ = \sum_{j\in S} \overline{v}_\alpha(j)$ in (3,7) and (3.8), we get that

(3.9)
$$\underline{v}^\alpha(i) + \varphi_\alpha^- = \max_{a\in A}\{r(i,a) + \min_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^n q_{ia}(j)\underline{v}_\alpha(j)\} \ (i \in S),$$

(3.10)
$$\overline{v}^\alpha(i) + \varphi_\alpha^+ = \max_{a\in A}\{r(i,a) + \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^n q_{ia}(j)\overline{v}_\alpha(j)\} \ (i \in S).$$

It is easily shown that $\underline{v}_\alpha \geq \underline{v}_{\alpha'}$, $\overline{v}_\alpha \leq \overline{v}_{\alpha'}$ $(0 \leq \alpha' \leq \alpha \leq 1)$. Also we have that $\underline{v}_\alpha$ and $\overline{v}_\alpha$ are continuous from below in $\alpha \in [0,1]$ (cf. [4]). So, applying the representative theorem (cf. [4]), we can construct the fuzzy numbers $\widetilde{v}_i$ $(i \in S)$ and $\widetilde{\varphi}$ by

(3.11) $$\widetilde{v}_i(x) = \sup_{\alpha\in[0,1]}\{\alpha \wedge \mathbf{1}_{[\underline{v}_\alpha(i), \overline{v}_\alpha(i)]}(x)\} \ (x \in \mathbb{R}),$$

(3.12) $$\widetilde{\varphi}(x) = \sup_{\alpha\in[0,1]}\{\alpha \wedge \mathbf{1}_{[\varphi_\alpha^-(i), \varphi_\alpha^+]}(x)\} \ (x \in \mathbb{R}).$$

Then, $\widetilde{\varphi}$ and $\widetilde{v}_i$ $(i \in S)$ satisfy (3.2). In fact, by (3.11) and (3.12), the α-cuts of $\widetilde{v}_i$ and $\widetilde{\varphi}$ are equal to $\widetilde{v}_{i\alpha} = [\underline{v}_\alpha(i), \overline{v}_\alpha(i)]$ and $\widetilde{\varphi}_\alpha = [\widetilde{\varphi}_\alpha^-, \widetilde{\varphi}_\alpha^+]$. So, the α-cut representation of (3.2) becomes (3.9) and (3.10). Also, the uniqueness of $\widetilde{\varphi}$ in (3.2) follows from the uniqueness of φ_α^- and φ_α^+ in (3.9) and (3.10). □

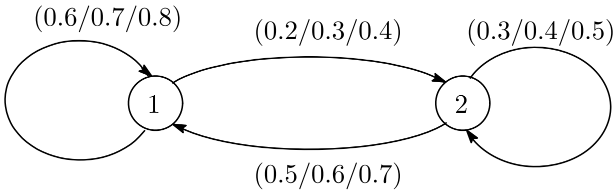
As a simple example, we consider a fuzzy perceptive model of a machine maintenance problem dealt with in ([11], p.17–18).

An example (a machine maintenance problem). A machine can be operated synchronously, say, once an hour. At each period there are two states; one is operating(state 1), and the other is in failure(state 2). If the machine fails, it can be restored to perfect functioning by repair. At each period, if the machine is running, we earn the return of \$ 3.00 per period; the fuzzy set of probability of being in state 1 at the next step is (0.6/0.7/0.8) and that of the probability of moving to state 2 is (0.2/0.3/0.4), where for any $0 \leq a < b < c \leq 1$, the triangular fuzzy number $(a/b/c)$ on $[0,1]$ is defined by

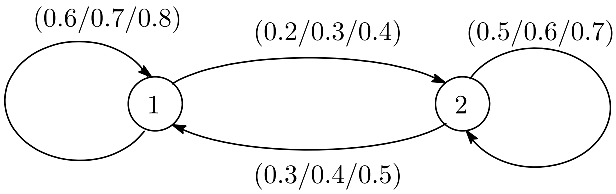
$$(a/b/c)(x) = \begin{cases} (x-a)/(b-a) \vee 0 & \text{if } 0 \leq x \leq b, \\ (x-c)/(b-c) \vee 0 & \text{if } b \leq x \leq 1. \end{cases}$$

If the machine is in failure, we have two actions to repair the failed machine; one is a rapid repair, denoted by 1, that yields the cost of \$ 2.00(that is, a return of −\$2.00) with the fuzzy set (0.5/0.6/0.7) of the probability moving in state 1 and the fuzzy set (0.3/0.4/0.5) of the probability being in state 2; another is a usual repair, denoted by 2, that requires the cost of \$1.00(that is, a return of −\$1.00) with the fuzzy set (0.3/0.4/0.5) of the probability moving in state 1 and the fuzzy set (0.5/0.6/0.7) of the probability being in state 2.

For the model considered, $S = \{1, 2\}$ and there exists two stationary policies, $F = \{f_1, f_2\}$ with $f_1(2) = 1$ and $f_2(2) = 2$, where f_1 denotes a policy of the usual repair and f_2 a policy of the rapid repair. The state transition diagrams of two policies are shown in Figure 1.



(a) Rapid repair f_1



(b) Usual repair f_2

Figure.1 Transition diagrams.

Using (2.3), we obtain $\widetilde{Q}_{ia}(\cdot)$ $(i \in S, a \in A)$, whose α-cut is given as follows(cf. [6]):

$$\begin{aligned}
\widetilde{Q}_{11,\alpha} &= co\{(.6 + .1\alpha,\ .4 - .1\alpha), (.8 - .1\alpha,\ .2 + .1\alpha)\},\\
\widetilde{Q}_{21,\alpha} &= co\{(.5 + .1\alpha,\ .5 - .1\alpha), (.7 - .1\alpha,\ .3 + .1\alpha)\},\\
\widetilde{Q}_{22,\alpha} &= co\{(.3 + .1\alpha,\ .7 - .1\alpha), (.5 - .1\alpha,\ .5 + .1\alpha)\},
\end{aligned}$$

where coX is a convex hull of a set X.

So, putting $x_1 = \widetilde{v}^-_{1\alpha}$, $x_2 = \widetilde{v}^+_{1\alpha}$, $\widetilde{v}^-_{2\alpha} = 0$, $\widetilde{v}^+_{2\alpha} = 0$, $y_1 = \widetilde{\varphi}^-_\alpha$, $y_2 = \widetilde{\varphi}^+_\alpha$, the α-cut optimality equations (3.3) and (3.10) become:

$$\begin{aligned}
x_1 + y_1 &= 3 + \min\{(.6 + .1\alpha)x_1,\ (.8 - .1\alpha)x_1\},\\
y_1 &= \max\big[-2 + \min\{(.5 + .1\alpha)x_1,\ (.7 - .1\alpha)x_1\},\\
&\qquad -1 + \min\{(.3 + .1\alpha)x_1,\ (.5 - .1\alpha)x_1\}\big],\\
x_2 + y_2 &= 3 + .9\max\{(.6 + .1\alpha)x_2,\ (.8 - .1\alpha)x_2\},\\
y_2 &= \max\big[-2 + \max\{(.5 + .1\alpha)x_2,\ (.7 - .1\alpha)x_2\},\\
&\qquad -1 + \max\{(.3 + .1\alpha)x_2,\ (.5 - .1\alpha)x_2\}\big],
\end{aligned}$$

After a simple calculation, we get

$$x_1 = x_2 = \frac{50}{9},\ y_1 = \frac{7}{9} + \frac{5}{9}\alpha,\ y_2 = \frac{17}{9} - \frac{5}{9}\alpha.$$

Thus, the average fuzzy perceptive value is

$$\widetilde{\varphi} = (\frac{7}{9}/\frac{12}{9}/\frac{17}{9}).$$

REFERENCES

[1] Blackwell,D., Discrete dynamic programming, *Ann. Math. Statist.*, **33**, (1962), 719–726.

[2] Dantzig,G.B., Folkman,J. and Shapiro,N., On the continuity of the minimum set of a continuous function, *J. Math. Anal. Appl.*, **17**, (1967), 519–548.

[3] Derman,C., *Finite State Markovian Decision Processes*, Academic Press, New York, (1970).

[4] Dubois,D. and Prade,H., *Fuzzy Sets and Systems : Theory and Applications*, Academic Press, (1980).

[5] Howard,R., *Dynamic Programming and Markov Process*, MIT Press, Cambridge, MA, (1960).

[6] Kurano,M., Song,J., Hosaka,M. and Huang,Y., Controlled Markov set-chains with discounting, *J. Appl. Prob.*, **35**, (1998), 293–302.

[7] Kurano,M., Yasuda,M. Nakagami,J. and Yoshida,Y., Ordering of fuzzy sets – A brief survey and new results, *J. Operations Research Society of Japan*, **43**, (2000), 138–148.

[8] Kurano,M., Yasuda,M. Nakagami,J. and Yoshida,Y., A fuzzy treatment of uncertain Markov decision process, RIMS Kokyuroku, Kyoto University, **1132**, (2000), 221–229.

[9] Kurano,M., Yasuda,M. Nakagami,J. and Yoshida,Y., A fuzzy stopping problem with the concept of perception, *Fuzzy Optimization and Decision Making*, **3**, (2004), 367–374.

[10] Kurano,M., Yasuda,M. Nakagami,J. and Yoshida,Y., Fuzzy perceptive values for MDPs with discounting, (Preprint).

[11] Mine,H. and Osaki,S., *Markov Decision Process*, Elsevier, Amsterdam, (1970).

[12] Nummelin,E., *General irreducible Markov chains and non-negative operators*, Cambridge University Press, (1984).

[13] Puterman,M.L., *Markov Decision Process: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, INC, (1994).

[14] Schweizer,D.T., Perturbation theory and finite Markov chains, *J. Applied Probab.*, **5**, (2068), 401–413.

[15] Solan,E., Continuity of the value of competitive Markov decision processes, *J. Theoretical Probability*, **16**, (2004), 831–845.

[16] Yoshida,Y. and Kerre,E.E., A fuzzy ordering on multi-dimensional fuzzy sets induced from convex cones, *Fuzzy Sets and Systems*, **130**, (2002), 343–355.

[17] Zadeh,L.A., Fuzzy sets, *Inform. and Control*, **8**, (1965), 338–353.

[18] Zadeh,L.A., Toward a perception-based theory of probabilistic reasoning with imprecise probabilities, *J. of Statistical Planning and Inference*, **105**, (2002), 233–264.