数理解析研究所講究録 1457 巻 2005 年 139-146

Fuzzy Perceptive Values for MDPs with Discounting

千葉大学教育学部 蔵野正美 (Masami Kurano)
Faculty of Education, Chiba University
千葉大学理学部 安田正實 (Masami Yasuda)
千葉大学理学部 中神潤一 (Jun-ichi Nakagami)
Faculty of Science, Chiba University
北九州大学経済学部 吉田祐治 (Yuji Yoshida)
Faculty of Economics and Business Administration, Kitakyushu University

Abstract

In this paper, we formulate the fuzzy perceptive model for discounted Markov decision processes in which the perception for transition probabilities is described by fuzzy sets. The optimal expected reward, called a fuzzy perceptive value, is characterized and calculated by a new fuzzy relation. As a numerical example, a machine maintenance problem is considered.
Keywords : Fuzzy perceptive model, Markov decision process,
fuzzy perceptive reward, optimal policy function.

1. Introduction and notation

Many contributions to Markov decision processes(MDPs) have been made (cf. [1], [2], [4], [9], [10]), in which the transition probability of the state at each time is assumed to be uniquely given. In a real application of MDPs, the transition probability will be estimated through the measurement of various phenomena. In such a case, the real value of the state transition probability may be partially observed by dimness of perception or measurement imprecision. For example, in a famous automobile replacement problem [4], the true value of the probability q_{ij} that the car in within age j after six months, given that the car is within age i at that time, may not be observed exactly. Usually, it is linguistically or roughly perceived, e.g., about 0.3, the probability considerably larger than 0.3, etc. A possible approach to handle such a case is to use the fuzzy set ([3], [12]), whose membership function can describe the perception value of the true probability. If the fuzzy perception of the transition probabilities for MDPs is given, how can we estimate in advance the future expected reward, called a fuzzy perceptive value, under the condition that we can know the true value of the transition probability immediately before our decision making.

In our previous work [8], we have tried the perceptive analysis for an optimal stopping problem. In this paper, we formulate the fuzzy perceptive model for MDPs and develop the perceptive analysis in which the fuzzy perceptive value for MDPs is characterized and calculated by a new fuzzy relation.

In remainder of this section, we will give some notation and fundamental results on MDPs, by which the fuzzy perceptive model is formulated in the sequel. For non-perception approaches to MDPs with fuzzy imprecision refer to [7]. Recently Zadeh [13] wrote a summary paper of perception based probability theory.

Let $\mathbb{R}, \mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ be the sets of real numbers, real n-dimensional column 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 $x = (x(1), x(2), \ldots, x(n))' \in \mathbb{R}^n$, $\|x\| = \sum_{j=1}^n |x(j)|$ and for $y = (y_{ij}) \in \mathbb{R}^{m\times n}$, $\|y\| = \max_{1\leq i\leq m} \sum_{j=1}^n |y_{ij}|$.

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 $\widetilde{x}_\alpha := \{x \in X \mid \widetilde{x}(x) \geq \alpha\}$ $(\alpha \in (0,1])$ 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 $x \leq y$ and $\widetilde{r}(x) \leq \widetilde{s}(y)$; (ii) for any $y \in \mathbb{R}$, there exists $x \in \mathbb{R}$ such that $x \leq y$ and $\widetilde{s}(y) \leq \widetilde{r}(x)$: Obviously, the binary relation $\preccurlyeq$ satisfies the axioms of a partial order relation on $\mathcal{F}(\mathbb{R})$ (cf. [6], [11]).

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. [6]): (i) $\widetilde{r} \preccurlyeq \widetilde{s}$; (ii) $\widetilde{r}_\alpha^- \leq \widetilde{s}_\alpha^-$ and $\widetilde{r}_\alpha^+ \leq \widetilde{s}_\alpha^+$ $(\alpha \in [0,1])$; (iii) $\widetilde{\max}\{\widetilde{r}, \widetilde{s}\} = \widetilde{s}$; (iv) $\widetilde{\min}\{\widetilde{r}, \widetilde{s}\} = \widetilde{r}$.

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 column vectors whose elements are in $\mathbb{C}_+$.

We have the following.

Lemma 1.1 ([5]) *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 x' denotes the transpose of a vector $x \in \mathbb{R}^n$ and for $g = (g_1, g_2, \ldots, g_n)' \in \mathbb{R}_+^n$ and $d = (d_1, d_2, \ldots, d_n)' \in D$, $g' \cdot d = \sum_{j=1}^n g_j d_j$.

Here, we define 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,

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

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 $j \in S$.

We wish to maximize the expected total discounted reward over the infinite horizon.

Denote by F the set of functions from S to A. A policy π is a sequence $(f_1, f_2, \dots)$ of functions with $f_t \in F$ $(t \geq 1)$. Let Π denote the class of policies. We denote by f^∞ the policy $(f_1, f_2, \dots)$ with $f_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary.

We associate with each $f \in F$ and $Q \in \mathcal{P}(S|SA)$ the column vector $r(f) = (r(1, f(1)), \dots, r(n, f(n)))'$ and the $n \times n$ transition matrix $Q(f)$, whose (i,j) element is $q_{i,f(i)}(j)$ $1 \leq i, j \leq n$. Then, the expected total discounted reward from $\pi = (f_1, f_2, \dots)$ is the column vector $\psi(\pi|Q) = (\psi(1, \pi|Q), \dots, \psi(n, \pi|Q))'$, which is defined, as a function of $Q \in \mathcal{P}(S|SA)$, by

$$\psi(\pi|Q) = \sum_{t=0}^{\infty} \beta^t Q(f_1) Q(f_2) \cdots Q(f_t) r(f_{t+1}), \tag{1.1}$$

where $0 < \beta < 1$ is a discount factor.

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

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

is said to be Q-optimal, and $\psi(Q) := (\psi(1|Q), \psi(2|Q), \cdots, \psi(n|Q))'$ is called the Q-optimal value vector.

We can state the well-known results.

Theorem 1.1 (cf. [2],[9],[10]) *For any $Q = (q_{ia}(\cdot) : i \in S, a \in A) \in \mathcal{P}(S|SA)$, the following holds:*

(i) *The Q-optimal value vector $\psi(Q) := (\psi(1|Q), \psi(2|Q), \cdots, \psi(n|Q))'$ is a unique solution to the optimality equations*

$$\psi(i|Q) = \max_{a \in A} \{r(i,a) + \beta \sum_{j \in S} q_{ia}(j) \psi(j|Q)\} \quad (i \in S); \tag{1.2}$$

(ii) *There exits an optimal stationary policy f_*^∞ such that $f_*(i) \in A$ attains the minimum in (1.2), i.e.,*

$$\psi(i|Q) = r(i, f_*(i)) + \beta \sum_{j \in S} q_{if_*(i)}(j) \psi(j|Q) \quad (i \in S). \tag{1.3}$$

In Section 2, we define a fuzzy-perceptive model for MDPs, which is analyzed in Section 3 with a numerical example. The proof of the theorem is given in Section 4.

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 $q_{ia}(\cdot) = (q_{ia}(1), q_{ia}(2), \dots, q_{ia}(n))'$, $\widetilde{Q}_{ia}(\cdot)$, 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:

$$\widetilde{Q}(Q) = \min_{i \in S, a \in A} \widetilde{Q}_{ia}(q_{ia}(\cdot)), \quad \text{where} \quad Q = (q_{ia}(\cdot) : i \in S, a \in A) \in \mathcal{P}(S|SA). \tag{2.1}$$

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

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

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

$$\widetilde{Q}_{ia}(q) = \min_{j \in S} \widetilde{q}_{ia}(j)(q_{ia}(j)), \quad \text{where} \quad q = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n)) \in \mathcal{P}(S). \tag{2.3}$$

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 $\widetilde{\psi}$ is a fuzzy set on $\mathbb{R}$ denoted by

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

The policy function $\delta^* \in \Delta$ is said to be optimal if $\widetilde{\psi}(i,\delta) \preccurlyeq \widetilde{\psi}(i,\delta^*)$ 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 $\widetilde{\psi} = (\widetilde{\psi}(1), \widetilde{\psi}(2), \ldots, \widetilde{\psi}(n))$ will be called a fuzzy perceptive value, where $\widetilde{\psi}(i) = \widetilde{\psi}(i,\delta^*)$ $(i \in S)$.

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 $\widetilde{\psi}(i,\delta)$ to be a fuzzy number is given in the following lemma.

Lemma 3.1 *For any $\delta \in \Delta$, if $\psi(i,\delta|Q)$ is continuous in $Q \in \mathcal{P}(S|SA)$, then $\widetilde{\psi}(i,\delta) \in \widetilde{\mathbb{R}}$.*

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

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

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

Theorem 1.1 in Section 1 guarantees that for each $Q \in \mathcal{P}(S|SA)$ 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.

Lemma 3.2 $\widetilde{\psi}(i,\delta^*) \in \widetilde{\mathbb{R}}$ *for all $i \in S$.*

Proof. Applying Lemma 3.1, it is sufficient to prove that $\widetilde{\psi}(i,\delta^*|Q)$ is continuous in $Q \in \mathcal{P}(S|SA)$. For simplicity, for any $Q \in \mathcal{P}(S|SA)$, we put $\psi(Q) = (\psi_1(Q), \psi_2(Q), \ldots, \psi_n(Q))'$ where $\psi_i(Q) = \psi(i,\delta^*|Q)$ $(i \in S)$. Let $Q = (q_{ia}(\cdot))$, $\overline{Q} = (\overline{q}_{ia}(\cdot)) \in \mathcal{P}(S|SA)$. By Theorem 1.1, we have:

$$\psi_i(Q) = \max_{a\in A}\{r(i,a) + \beta \sum_{j\in S} q_{ia}(j)\psi_j(Q)\}; \tag{3.1}$$

$$\psi_i(\overline{Q}) = \max_{a\in A}\{r(i,a) + \beta \sum_{j\in S} \overline{q}_{ia}(j)\psi_j(\overline{Q})\}. \tag{3.2}$$

Suppose that $a_i = \delta^*(Q)(i)$ and $\overline{a}_i = \delta^*(\overline{Q})(i)$ $(i \in S)$ give the minimum in (3.1) and (3.2) respectively. Let $a = (a_1, a_2, \ldots, a_n)'$ and $\overline{a} = (\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_n)'$. Then, it yields that

$$\begin{aligned}
\psi(Q) - \psi(\overline{Q}) &\leq (r(a) + \beta Q(a)\psi(Q)) - (r(a) + \beta\overline{Q}(a)\psi(\overline{Q})) \\
&= \beta(Q(a)\psi(Q) - \overline{Q}\psi(\overline{Q})) \\
&= \beta(Q(a) - \overline{Q}(a))\psi(Q) + \beta\overline{Q}(a)(\psi(Q) - \psi(\overline{Q})),
\end{aligned}$$

where $r(a) = (r(1,a_1), r(2,a_2), \ldots, r(n,a_n))'$ and $Q(a) = (q_{ia_i}(j))$ and $\overline{Q}(a) = (\overline{q}_{ia_i}(j))$.

Thus, we get

$$(I - \beta\overline{Q}(a))(\psi(Q) - \psi(\overline{Q})) \leq \beta(Q(a) - \overline{Q}(a))\psi(Q),$$

where I is an identity matrix. Since $(I - \beta\overline{Q}(a))^{-1} = \sum_{k=0}^{\infty} \beta^k \overline{Q}(a)^k \geq 0$, we have

$$\psi(Q) - \psi(\overline{Q})) \leq \beta(I - \beta\overline{Q}(a))^{-1}(Q(a) - \overline{Q}(a))\psi(\overline{Q}), \tag{3.3}$$

Similarly we get

$$\psi(\overline{Q}) - \psi(Q)) \leq \beta(I - \beta Q(\overline{a}))^{-1}(\overline{Q}(\overline{a}) - Q(\overline{a}))\psi(Q), \tag{3.4}$$

where $Q(\overline{a})$ and $\overline{Q}(\overline{a})$ are defined similarly as the above. Observing that

$$0 \leq \psi_i(Q), \psi_i(\overline{Q}) \leq \frac{1}{1-\beta} \max_{i\in S, a\in A} r(i,a) =: M$$

and $\|Q(a) - \overline{Q}(a)\| \leq \|Q - \overline{Q}\|$ and $\|Q(\overline{a}) - \overline{Q}(\overline{a})\| \leq \|Q - \overline{Q}\|$, from (3.3) and (3.4), it holds that

$$\|\psi(Q) - \psi(\overline{Q})\| \leq \beta M \max\{\|(I - \beta\overline{Q}(a))^{-1}\|, \|(I - \beta Q(\overline{a}))^{-1}\|\} \cdot \|Q - \overline{Q}\|. \tag{3.5}$$

When $Q \to \overline{Q}$ in $\mathcal{P}(S|SA)$, $\|(I-\beta\overline{Q}(a))^{-1}\|$ and $\|(I-\beta Q(\bar{a}))^{-1}\|$ are bounded and (3.5) means that $\|\psi(Q)-\psi(\overline{Q})\| \to 0$. □

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

$$\psi(i,\delta|Q) \le \psi(i,\delta^*|Q) \quad (i \in S). \tag{3.6}$$

For any $x \in \mathbb{R}$, let $\alpha := \widetilde{\psi}(i,\delta)(x)$. Then, from the definition there exists $Q \in \widetilde{Q}_\alpha$ with $x = \psi(i,\delta|Q)$. By (3,6), $y := \psi(i,\delta^*|Q) \ge x$, which implies $\widetilde{\psi}(i,\delta^*)(y) \ge \alpha$.

On the other hand, for $y \in \mathbb{R}$, let $\alpha := \widetilde{\psi}(i,\delta^*)(y)$. Then, there exists $Q \in \widetilde{Q}_\alpha$ such that $y = \psi(i,\delta^*|Q)$. From (3.6), we have that $y \ge x := \psi(i,\delta|Q)$. This implies $\widetilde{\psi}(i,\delta|Q) \le \alpha$. The above discussion yields that $\widetilde{\psi}(i,\delta) \preccurlyeq \widetilde{\psi}(i,\delta^*)$. □

From Lemma 3.2, $\widetilde{\psi}(i) := \widetilde{\psi}(i,\delta^*) \in \widetilde{\mathbb{R}}$ $(i \in S)$, so that we denote by $\widetilde{\psi}_\alpha(i) := [\widetilde{\psi}^-_\alpha(i), \widetilde{\psi}^+_\alpha(i)]$, the α-cut of $\widetilde{\psi}(i)$.

The fuzzy perceptive value $\widetilde{\psi} = (\widetilde{\psi}(1), \ldots, \widetilde{\psi}(n))'$ is characterized by a new fuzzy optimality relation in Theorem 3.2, whose proof is omitted.

Theorem 3.2 *The fuzzy perceptive value $\widetilde{\psi} = (\widetilde{\psi}(1), \widetilde{\psi}(2), \ldots, \widetilde{\psi}(n))'$ is a unique solution to the following fuzzy optimality relations:*

$$\widetilde{\psi}(i) = \widetilde{\max_{a \in A}}\{1_{\{r(i,a)\}} + \beta \widetilde{Q}_{ia} \cdot \widetilde{\psi}\} \ (i \in S), \tag{3.7}$$

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

The α-cut expression of (3.7) is as follows:

$$\widetilde{\psi}^-_\alpha(i) = \max_{a \in A}\{r(i,a) + \beta \min_{q_{ia} \in \widetilde{Q}_{ia,\alpha}} q_{ia} \cdot \widetilde{\psi}^-_\alpha\} \quad (i \in S); \tag{3.8}$$

$$\widetilde{\psi}^+_\alpha(i) = \max_{a \in A}\{r(i,a) + \beta \max_{q_{ia} \in \widetilde{Q}_{ia,\alpha}} q_{ia} \cdot \widetilde{\psi}^+_\alpha\} \quad (i \in S), \tag{3.9}$$

where $\widetilde{\psi}^\mp_\alpha = (\widetilde{\psi}^\mp(1), \widetilde{\psi}^\mp(2), \ldots, \widetilde{\psi}^\mp(n))'$ and $q_{ia} \cdot \widetilde{\psi}^\mp_\alpha = \sum_{j \in S} q_{ia}(j)\widetilde{\psi}^\mp_\alpha(j)$.

We note that the α-cut of $\widetilde{Q}_{ia} \cdot \widetilde{\psi}$ in (3.7) is in $\mathbb{C}$ from Lemma 1.1, so that $\widetilde{Q}_{ia} \cdot \widetilde{\psi} \in \widetilde{\mathbb{R}}$. Thus, the right hand of (3.7) is well-defined.

As a simple example, we consider a fuzzy perceptive model of a machine maintenance problem dealt with in ([9], p.1, p.17–18).

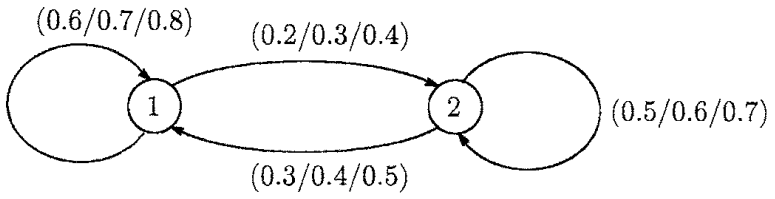
An example (a machine maintenance problem). A machine can be operated synchronously, say, once an hour. At each period there are two states; one is operating(state 1), and the other is in failure(state 2). If the machine fails, it can be restored to perfect functioning by repair. At each period, if the 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 $0 \le a < b < c \le 1$, the fuzzy number $(a/b/c)$ on $[0,1]$ is defined by

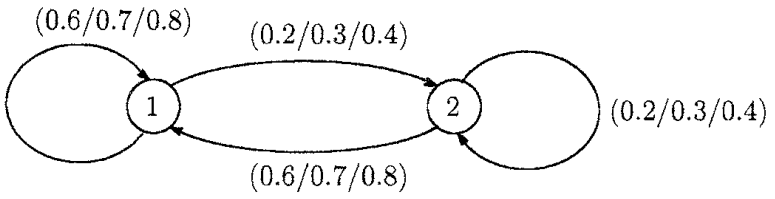
$$(a/b/c)(x) = \begin{cases} (x-a)/(b-a) \vee 0 & \text{if } 0 \le x \le b, \\ (x-c)/(b-c) \vee 0 & \text{if } b \le x \le 1. \end{cases}$$

If the machine is in failure, we have two actions to repair the failed machine; one is a usual repair, denoted by 1, that yields 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; another is a rapid repair, denoted by 2, that requires the cost of \$2.00(that is, a return of −\$2.00) with the fuzzy set (0.6/0.7/0.8) of the probability moving in state 1 and the fuzzy set (0.2/0.3/0.4) 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) Usual repair f_1



(b) Rapid 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. [5]):

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

where coX is a convex hull of a set X.

So, putting $x_1 = \widetilde{\psi}_\alpha^-(1)$, $x_2 = \widetilde{\psi}_\alpha^-(2)$, $y_1 = \widetilde{\psi}_\alpha^+(1)$, $y_2 = \widetilde{\psi}_\alpha^+(2)$, the α-cut optimality equations (3.8) and (3.9) with $\beta = 0.9$ become:

$$\begin{aligned} x_1 &= 3 + .9\min\{(.6+.1\alpha)x_1 + (.4-.1\alpha)x_2,\ (.8-.1\alpha)x_1 + (.2+.1\alpha)x_2\} \\ x_2 &= \max\big[-1 + .9\min\{(.3+.1\alpha)x_1 + (.7-.1\alpha)x_2,\ (.5-.1\alpha)x_1 + (.5+.1\alpha)x_2\}, \\ &\qquad -2 + .9\min\{(.6+.1\alpha)x_1 + (.4-.1\alpha)x_2,\ (.8-0.1\alpha)x_1 + (.2+.1\alpha)x_2\}\big], \\ y_1 &= 3 + .9\max\{(.6+.1\alpha)y_1 + (.4-.1\alpha)y_2,\ (.8-.1\alpha)y_1 + (.2+.1\alpha)y_2\} \\ y_2 &= \max\big[-1 + .9\max\{(.3+.1\alpha)y_1 + (.7-.1\alpha)y_2,\ (.5-.1\alpha)y_1 + (.5+.1\alpha)y_2\}, \\ &\qquad -2 + .9\max\{(.6+.1\alpha)y_1 + (.4-.1\alpha)y_2,\ (.8-0.1\alpha)y_1 + (.2+.1\alpha)y_2\}\big], \end{aligned}$$

After a simple calculation, we get

$$x_1 = 12 + 4.5\alpha,\ x_2 = 7 + 4.5\alpha,\ y_1 = 21 - 4.5\alpha,\ y_2 = 16 - 4.5\alpha.$$

Thus, we know the fuzzy perceptive value is

$$\widetilde{\psi}(1) = (12/16.5/21),\ \ \widetilde{\psi}(2) = (7/11.5/16).$$

References

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

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

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

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

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

[6] 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.

[7] Kurano,M., Yasuda,M. Nakagami,J. and Yoshida,Y., A fuzzy treatment of uncertain Markov decision process, 数理解析研究所講究録, **1132**, (2000), 221–229.

[8] 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.

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

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

[11] 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.

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

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