Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Fuzzy Sets and Systems ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮
FUZZY sets and systems
www.elsevier.com/locate/fss

Fuzzy optimality relation for perceptive MDPs—the average case

M. Kurano[a,*], M. Yasuda[b], J. Nakagami[b], Y. Yoshida[c]

[a] Faculty of Education, Chiba University, Chiba 263-8522, Japan
[b] Faculty of Science, Chiba University, Chiba 263-8522, Japan
[c] Faculty of Economics & Business Administration, Kitakyushu University, Kitakyushu 802-8577 Japan

Abstract

In this paper, the fuzzy perceptive model for average reward Markov decision processes is defined and a method of computing the corresponding fuzzy perceptive values is proposed. Under the minorization condition for fuzzy perceptive transition matrices, it is characterized by the optimal average expected reward, called the average perceptive value, using a fuzzy optimality relation. Also, we give a simple numerical example.

Keywords: Fuzzy perceptive model; Markov decision process; Average criterion; Fuzzy perceptive value; Optimal policy function

1. Introduction

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 fuzzy sets (cf. [4,19]), whose membership function describes the level of the perception of the required data. And refer to [15,14] for a possibility fuzzy theory in uncertain MDPs. 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 [20] using a generalized extension principle.

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 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. 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].

* Corresponding author. Tel.: +81 43 290 2669.
E-mail addresses: kurano@faculty.chiba-u.jp (M. Kurano), yasuda@math.s.chiba-u.ac.jp (M. Yasuda), nakagami@math.s.chiba-u.ac.jp (J. Nakagami), yoshida@kitakyu-u.ac.jp (Y. Yoshida).

0165-0114/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.fss.2007.04.006

2. Notations

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 we put $\|x\| = \sum_{j=1}^{n} |x(j)|$ for a vector $x = (x(1), x(2), \ldots, x(n)) \in \mathbb{R}^n$ and we write $\|y\| = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |y_{ij}|$ for a matrix $y = (y_{ij}) \in \mathbb{R}^{m\times n}$. For any set X, let $\mathcal{F}(X)$ be the set of all fuzzy sets $\widetilde{x} : X \mapsto [0, 1]$. The α-cut of $\widetilde{x} \in \mathcal{F}(X)$ is given by $\widetilde{x}_\alpha := \{x \in X | \widetilde{x}(x) \geqslant \alpha\} (\alpha \in (0, 1])$ and $\widetilde{x}_0 := \mathrm{cl}\{x \in X | \widetilde{x}(x) > 0\}$, where cl is the 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 $\widetilde{r} \in \mathcal{F}(\mathbb{R})$ and $\widetilde{r}$ is normal, upper semi-continuous 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 $\widetilde{r}$ and $\widetilde{s}$ satisfy the following (i) and (ii): (i) for any $x \in \mathbb{R}$, there exists $y \in \mathbb{R}$ such that $x \leqslant y$ and $\widetilde{r}(x) \leqslant \widetilde{s}(y)$; (ii) for any $y \in \mathbb{R}$, there exists $x \in \mathbb{R}$ such that $x \leqslant y$ and $\widetilde{s}(y) \leqslant \widetilde{r}(x)$. Obviously, the binary relation $\preccurlyeq$ satisfies the axioms of a partial order relation on $\mathcal{F}(\mathbb{R})$ (cf. [7,18]).

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

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

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

respectively, where $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for any $a, b \in \mathbb{R}$. It is easily proved that $\widetilde{\max}\{\widetilde{r}, \widetilde{s}\} \in \widetilde{\mathbb{R}}$ and $\widetilde{\min}\{\widetilde{r}, \widetilde{s}\} \in \widetilde{\mathbb{R}}$ for $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$. It is known that the following (i)–(iv) are equivalent each other (cf. [7]): (i) $\widetilde{r} \preccurlyeq \widetilde{s}$; (ii) $\widetilde{r}_\alpha^- \leqslant \widetilde{s}_\alpha^-$ and $\widetilde{r}_\alpha^+ \leqslant \widetilde{s}_\alpha^+$ $(\alpha \in [0, 1])$; (iii) $\widetilde{\max}\{\widetilde{r}, \widetilde{s}\} = \widetilde{s}$; (iv) $\widetilde{\min}\{\widetilde{r}, \widetilde{s}\} = \widetilde{r}$. Also we use the addition by $(\widetilde{r} + \widetilde{s})(y) := \sup_{\substack{x_1, x_2 \in \mathbb{R} \\ y = x_1 + x_2}} \{\widetilde{r}(x_1) \wedge \widetilde{s}(x_2)\} (y \in \mathbb{R})$ for any $\widetilde{r}, \widetilde{s} \in \widetilde{\mathbb{R}}$. 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 let $\mathbb{C}_+^n$ the set of all n-dimensional vectors whose elements are in $\mathbb{C}_+$.

Lemma 2.1 (*Kurano et al. [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*

$$GD = \left\{ g \cdot d = \sum_{j=1}^{n} g_j d_j \,\middle|\, g \in G, d \in D \right\} \in \mathbb{C}_+$$

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 3. Consider a finite state space $S = \{1, 2, \ldots, n\}$ and a finite action space $A = \{1, 2, \ldots, k\}$, where n and k are fixed positive integers. 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 when an elements of $S \times A$ is given, that is,

$$\mathcal{P}(S) := \left\{ q = (q(1), q(2), \ldots, q(n)) \,\middle|\, q(i) \geqslant 0, \sum_{i=1}^{n} q(i) = 1, i \in S \right\},$$

$$\mathcal{P}(S|SA) := \{Q = (q_{ia} : i \in S, a \in A) | q_{ia} = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n)) \in \mathcal{P}(S), i \in S, a \in A\}.$$

For any $Q = (q_{ia}) \in \mathcal{P}(S|SA)$, we define a controlled dynamic system $\mathcal{M}(Q)$, called an MDP, specified by $\{S, A, Q, r\}$, where $r : S \times A \mapsto \mathbb{R}_+$ is an immediate reward function. When the system is in state $i \in S$ and action $a \in A$ is taken, the system moves to a new state $j \in S$ selected according to $q_{ia}(j)$ and a reward $r(i, a)$ is

Please cite this article as: M. Kurano, et al., Fuzzy optimality relation for perceptive MDPs—the average case, Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss.2007.04.006

obtained. And at the next step the process goes on from the new state $j \in S$. The sample space is the product space $\Omega = (S \times A)^\infty$, and the projections $X_t : \Omega \mapsto S$ and $\Delta_t : \Omega \mapsto A$ describe a state and an action at time t, respectively ($t \geqslant 0$). A policy $\pi = (\pi_1, \pi_2, \ldots)$ is a sequence of conditional probabilities $\pi_t(\cdot|x_0, a_0, \ldots, x_t)$ on A 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 $\gamma = (\gamma(\cdot|i), i \in S)$ for which $\pi(\cdot|x_0, a_0, \ldots, x_t) = \gamma(\cdot|x_t)$ for all $t \geqslant 0$ 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 $i \in S$. For each $\pi \in \Pi$, an initial state $X_0 = i$ and a transition matrix $Q \in \mathcal{P}(S|SA)$, the probability measure $P_\pi(\cdot|X_0 = i, Q)$ on Ω is defined in a 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

$$\varphi(i, \pi|Q) = \liminf_{T\to\infty} \frac{1}{T} E_\pi(\varphi_T|X_0 = i, Q) \tag{2.1}$$

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

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

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

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 $Q = (q_{ia} : i \in S, a \in A) \in \mathcal{P}(S|SA)$ 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 $u : S \mapsto \mathbb{R}$. For any $Q = (q_{ia} : i \in S, a \in A) \in \mathcal{P}(S|SA)$, we define a map $U\{Q\} : \mathcal{B}(S) \mapsto \mathcal{B}(S)$ by

$$U\{Q\}u(i) := \max_{a\in A}\{r(i, a) + \sum_{j\in S}(q_{ia}(j) - \delta(Q))u(j) \tag{2.2}$$

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

$$U\{Q\}v = v. \tag{2.3}$$

Putting $\varphi(Q) = \delta(Q)\sum_{j\in S} v(j)$ in (2.3), we obtain an optimality equation for the average expected reward:

$$v(Q)(i) + \varphi(Q) = \max_{a\in A}\left\{r(i, a) + \sum_{j\in S} q_{ia}(j)v(Q)(j)\right\}. \tag{2.4}$$

The following lemma follows from (2.4). Refer to [1,3,5,13] as for the theory of MDPs.

Lemma 2.2. *Suppose that $Q \in \mathcal{P}(S|SA)$ satisfies Condition M. If $f(i) \in A^*(i|Q)$ for each $i \in S$ and $\varphi(i|Q)$ is independent of $i \in S$, and hence we put $\varphi(Q) := \varphi(i|Q)$, then f is Q-optimal where $A^*(i|Q) := \{a \in A | a$ maximizes the right-hand side of* (2.4)$\}$.

Let $\mathcal{P}_{\mathrm{M}}$ 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 2.3 (*cf. Schweizer [16], Solan [17]*). *The optimal average reward $\varphi(Q)$ is continuous in $\mathcal{P}_{\mathrm{M}}$.*

In Section 3, we define a fuzzy perceptive model for average reward MDPs, which is analyzed in Section 5 with a numerical example.

Please cite this article as: M. Kurano, et al., Fuzzy optimality relation for perceptive MDPs—the average case, Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss.2007.04.006

3. 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 a fuzzy set on $\mathcal{P}(S|SA)$ whose membership function $\widetilde{Q}$ describes a perception value of the transition probability.

Firstly, for each $i \in S$ and $a \in A$, we give a fuzzy perception of $q_{ia} = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n))$. Denote by $\widetilde{Q}_{ia}(\cdot)$ a fuzzy set on $\mathcal{P}(S)$ satisfying the following conditions (i) and (ii). (i) *Normality*: There exists a $q = q_{ia} \in \mathcal{P}(S)$ with $\widetilde{Q}_{ia}(q) = 1$. (ii) *Convexity and compactness*: For each $\alpha \in [0, 1]$, its α-cut $\widetilde{Q}_{ia,\alpha} = \{q = q_{ia} \in \mathcal{P}(S) | \widetilde{Q}_{ia}(q) \geqslant \alpha\}$ is a convex and compact subset in $\mathcal{P}(S)$.

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 $\mathcal{P}(S|SA)$, which is called fuzzy perception of the transition probability Q in MDPs, as follows:

$$\widetilde{Q}(Q) = \min_{i \in S, a \in A} \widetilde{Q}_{ia}(q_{ia}(\cdot)), \tag{3.1}$$

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

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

$$\begin{aligned}\widetilde{Q}_\alpha &= \{Q = (q_{ia} : i \in S, a \in A) \in \mathcal{P}(S|SA) | q_{ia} \in \widetilde{Q}_{ia,\alpha} \text{ for } i \in S, a \in A\} \\ &= \prod_{i \in S, a \in A} \widetilde{Q}_{ia,\alpha} \quad (\alpha \in [0, 1]).\end{aligned} \tag{3.2}$$

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 a fuzzy set $\widetilde{q}_{ia}(j) \in \widetilde{\mathbb{R}}(j \in S)$, which represents the fuzzy perception of $q_{ia}(j)$ (the transition probability to $j \in S$ when an action $a \in A$ is taken in state $i \in S$). Then, $\widetilde{Q}_{ia}(\cdot)$ is defined by

$$\widetilde{Q}_{ia}(q) = \min_{j \in S} \widetilde{q}_{ia}(j)(q_{ia}(j)), \tag{3.3}$$

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

$$\widetilde{\varphi}(i, \delta)(x) = \sup_{\substack{Q \in \mathcal{P}(S|PS) \\ x = \varphi(i, \delta(Q)|Q)}} \widetilde{Q}(Q) \quad (i \in S). \tag{3.4}$$

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 an optimal policy function δ^* and to characterize the fuzzy perceptive value.

4. Perceptive analysis

In this section, we derive a new fuzzy optimality relation to solve 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 4.1. *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 the 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

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

Please cite this article as: M. Kurano, et al., Fuzzy optimality relation for perceptive MDPs—the average case, Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss.2007.04.006

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

Lemma 2.2 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 the minorization condition for the perceptive model $\mathcal{M}(\widetilde{Q})$. We say that $\widetilde{Q}$ on $\mathcal{P}(S|SA)$ satisfies Condition M if $\widetilde{Q}_0 \subset \mathcal{P}_{\mathrm{M}}$, where $\widetilde{Q}_0$ is the 0-cut of $\widetilde{Q}$.

Lemma 4.2. *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 2.2, $\widetilde{\varphi}(i, \delta^*|Q)$ is continuous in $\widetilde{Q}_0$, so that $\widetilde{\varphi}(i, \delta^*) \in \widetilde{\mathbb{R}}$ follows from Lemma 4.1. Also, from Lemma 2.1, $\varphi(i, \delta^*)$ is clearly independent of $i \in S$. $\square$

Theorem 4.3. *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

$$\varphi(i, \delta|Q) \leqslant \varphi(i, \delta^*|Q) \quad (i \in S). \tag{4.1}$$

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), $y := \varphi(i, \delta^*|Q) \geqslant x$, which implies $\widetilde{\varphi}(i, \delta^*)(y) \geqslant \alpha$. 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 $y \geqslant x := \varphi(i, \delta|Q)$. This implies $\widetilde{\varphi}(i, \delta|Q) \leqslant \alpha$. The above discussion yields that $\widetilde{\varphi}(i, \delta) \preccurlyeq \widetilde{\varphi}(i, \delta^*)$. $\square$

From Lemma 4.2, we denote by $\widetilde{\varphi}_\alpha := [\widetilde{\varphi}_\alpha^-, \widetilde{\varphi}_\alpha^+] \in \mathbb{C}$ the α-cut of $\widetilde{\varphi} \in \widetilde{\mathbb{R}}(i \in S)$. In the following theorem, the fuzzy perceptive value $\widetilde{\varphi}$ is characterized by a fuzzy optimality relation.

Theorem 4.4. *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*:

$$\widetilde{v}_i + \widetilde{\varphi} = \widetilde{\max_{a \in A}}\{1_{\{r(i,a)\}} + \widetilde{Q}_{ia} \cdot \widetilde{v}\}, \tag{4.2}$$

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 $\{(q, \varphi)|x = \sum_{j=1}^n q(j)\varphi(j), q \in \mathcal{P}(S), \varphi \in \mathbb{R}^n)\}$ and $\widetilde{v}(\varphi) = \widetilde{v}_1(\varphi(1)) \wedge \cdots \wedge \widetilde{v}_n(\varphi(n))$.

The explicit form for the α-cut expression of (4.2) means as follows:

$$\widetilde{v}_{i,\alpha}^- + \widetilde{\varphi}^- = \max_{a \in A}\left\{r(i,a) + \min_{q_{ia} \in \widetilde{Q}_{ia,\alpha}} q_{ia} \cdot \widetilde{v}_\alpha^-\right\} \quad (\alpha \in [0,1]), \tag{4.3}$$

$$\widetilde{v}_{i,\alpha}^+ + \widetilde{\varphi}^+ = \max_{a \in A}\left\{r(i,a) + \max_{q_{ia} \in \widetilde{Q}_{ia,\alpha}} q_{ia} \cdot \widetilde{v}_\alpha^+\right\} \quad (\alpha \in [0,1]), \tag{4.4}$$

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 (4.2) is in $\mathbb{C}$ from Lemma 2.1, so that $\widetilde{Q}_{ia} \cdot \widetilde{v} \in \widetilde{\mathbb{R}}$. Thus, the right-hand side of (4.2) is well defined.

Proof. Under Condition M, we have $\widetilde{Q}_0 \subset \mathcal{P}_{\mathrm{M}}$, so that $\delta := \min_{Q \in \widetilde{Q}_0} \delta(Q) > 0$ and $q_{ia}(j) \geqslant \delta$ for all $q = (q_{ia}(\cdot)) \in \widetilde{Q}_{ia,\alpha}(\alpha \in [0,1])$. For any $\alpha \in [0,1]$, we define maps $\underline{U}^\alpha, \overline{U}^\alpha : \mathcal{B}(S) \mapsto \mathcal{B}(S)$ by

$$\underline{U}^\alpha u(i) = \min_{q_{ia} \in \widetilde{Q}_{ia,\alpha}} \max_{a \in A}\left\{r(i,a) + \sum_{j=1}^n (q_{ia}(j) - \delta)u(j)\right\} \quad (i \in S), \tag{4.5}$$

Please cite this article as: M. Kurano, et al., Fuzzy optimality relation for perceptive MDPs—the average case, Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss.2007.04.006

$$\overline{U}^{\alpha}u(i) = \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \max_{a\in A}\left\{r(i,a) + \sum_{j=1}^{n}(q_{ia}(j)-\delta)u(j)\right\} \quad (i\in S), \tag{4.6}$$

for any $u \in \mathcal{B}(S)$. Then, it is easily proved that the maps $\underline{U}^{\alpha}$ and $\overline{U}^{\alpha}$ are contractive with modulus $\beta = 1-\delta(<1)$. Thus, the unique fixed points exist for $\underline{U}^{\alpha}$ and $\overline{U}^{\alpha}$. Let us 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 **?**, we observe that $\underline{v}_{\alpha}$ and $\overline{v}_{\alpha}$ satisfy (4.7) and (4.8):

$$\underline{v}^{\alpha}(i) = \max_{a\in A}\left\{r(i,a) + \min_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^{n}(q_{ia}(j)-\delta)\underline{v}_{\alpha}(j)\right\} \quad (i\in S), \tag{4.7}$$

$$\overline{v}^{\alpha}(i) = \max_{a\in A}\left\{r(i,a) + \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^{n}(q_{ia}(j)-\delta)\overline{v}_{\alpha}(j)\right\} \quad (i\in S). \tag{4.8}$$

Putting $\varphi_{\alpha}^{-} = \sum_{j\in S}\underline{v}_{\alpha}(j)$ and $\varphi_{\alpha}^{+} = \sum_{j\in S}\overline{v}_{\alpha}(j)$ in (4.7) and (4.8), we get that

$$\underline{v}^{\alpha}(i) + \varphi_{\alpha}^{-} = \max_{a\in A}\left\{r(i,a) + \min_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^{n}q_{ia}(j)\underline{v}_{\alpha}(j)\right\} \quad (i\in S), \tag{4.9}$$

$$\overline{v}^{\alpha}(i) + \varphi_{\alpha}^{+} = \max_{a\in A}\left\{r(i,a) + \max_{q_{ia}\in\widetilde{Q}_{ia,\alpha}} \sum_{j=1}^{n}q_{ia}(j)\overline{v}_{\alpha}(j)\right\} \quad (i\in S). \tag{4.10}$$

It is easily shown that $\underline{v}_{\alpha} \geqslant \underline{v}_{\alpha'}$, $\overline{v}_{\alpha} \leqslant \overline{v}_{\alpha'} (0\leqslant\alpha'\leqslant\alpha\leqslant 1)$. 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 fuzzy numbers $\widetilde{v}_i (i\in S)$ and $\widetilde{\varphi}$ by

$$\widetilde{v}_i(x) = \sup_{\alpha\in[0,1]}\{\alpha \wedge \mathbf{1}_{[\underline{v}_{\alpha}(i),\overline{v}_{\alpha}(i)]}(x)\} \quad (x\in\mathbb{R}), \tag{4.11}$$

$$\widetilde{\varphi}(x) = \sup_{\alpha\in[0,1]}\{\alpha \wedge \mathbf{1}_{[\varphi_{\alpha}^{-}(i),\varphi_{\alpha}^{+}]}(x)\} \quad (x\in\mathbb{R}). \tag{4.12}$$

Then, $\widetilde{\varphi}$ and $\widetilde{v}_i (i\in S)$ satisfy (4.2). In fact, by (4.11) and (4.12), 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 (4.2) becomes (4.9) and (4.10). Also, the uniqueness of $\widetilde{\varphi}$ in (4.2) follows from the uniqueness of φ_{α}^{-} and φ_{α}^{+} in (4.9) and (4.10). □

5. An example for a machine maintenance

As a simple example, we consider a fuzzy perceptive model of a machine maintenance problem dealt with in [11], pp. 17–18. We consider a machine which is 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), the triangular fuzzy number $(a/b/c)$ on [0, 1] is defined by

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

where for any $0\leqslant a < b < c\leqslant 1$. 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

Please cite this article as: M. Kurano, et al., Fuzzy optimality relation for perceptive MDPs—the average case, Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss.2007.04.006

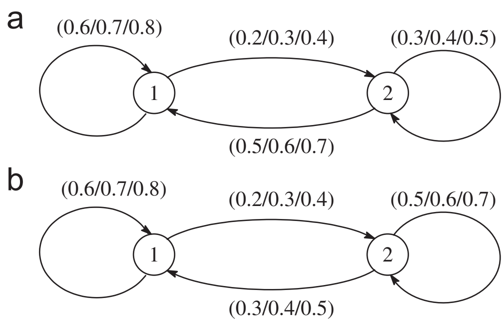

Fig. 1. (a) Transition diagram of the rapid repair f_1. (b) Transition diagram of the usual repair f_2.

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 Fig. 1(a) and (b).

Using (3.3), we obtain $\widetilde{Q}_{ia}(\cdot)(i \in S, a \in A)$, whose α-cut is given as follows (cf. [6]):

$$\widetilde{Q}_{11,\alpha} = co\{(.6 + .1\alpha, .4 - .1\alpha), (.8 - .1\alpha, .2 + .1\alpha)\},$$

$$\widetilde{Q}_{21,\alpha} = co\{(.5 + .1\alpha, .5 - .1\alpha), (.7 - .1\alpha, .3 + .1\alpha)\},$$

$$\widetilde{Q}_{22,\alpha} = co\{(.3 + .1\alpha, .7 - .1\alpha), (.5 - .1\alpha, .5 + .1\alpha)\},$$

where $co\{A, B\}$ is a convex hull of $A \cup B$.

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-cuts of the optimality Eqs. (4.3) and (4.10) become

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

After a simple calculation, we get

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

Thus, the average fuzzy perceptive value is a triangular fuzzy number

$$\widetilde{\varphi} = \left(\frac{7}{9}\Big/\frac{12}{9}\Big/\frac{17}{9}\right) = (0.778/1.333/1.889).$$

6. Conclusions with summary

We considered the average case of MDPs with fuzzy perceptive transition matrices. Not an optimality equation but a fuzzy optimality “relation" is defined, so that we could characterize the optimal average expected reward, that is, the average perceptive value. Also a numerical example for a machine maintenance is calculated under a triangular fuzzy number. The corresponding fuzzy relation is not so simple in general, because of innovating the fuzzy perceptive reward. It should be solved by cooperating an interval analysis method by expanding the usual Dynamic Programming.

Acknowledgment

The authors should express their thanks to two anonymous referees for the indication of typographical errors and useful comments. Also to the editors for their helpful suggestions.

References

[1] D. Blackwell, Discrete dynamic programming, Ann. Math. Statist. 33 (1962) 719–726.
[2] G.B. Dantzig, J. Folkman, N. Shapiro, On the continuity of the minimum set of a continuous function, J. Math. Anal. Appl. 17 (1967) 519–548.
[3] C. Derman, Finite State Markovian Decision Processes, Academic Press, New York, 1970.
[4] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.
[5] R. Howard, Dynamic Programming and Markov Process, MIT Press, Cambridge, MA, 1960.
[6] M. Kurano, J. Song, M. Hosaka, Y. Huang, Controlled Markov set-chains with discounting, J. Appl. Probab. 35 (1998) 293–302.
[7] M. Kurano, M. Yasuda, J. Nakagami, Y. Yoshida, Ordering of fuzzy sets—a brief survey and new results, J. Oper. Res. Soc. Japan 43 (2000) 138–148.
[8] M. Kurano, M. Yasuda, J. Nakagami, Y. Yoshida, A Fuzzy Treatment of Uncertain Markov Decision Process, vol. 1132, RIMS Kokyuroku, Kyoto University, 2000 pp. 221-229.
[9] M. Kurano, M. Yasuda, J. Nakagami, Y. Yoshida, A fuzzy stopping problem with the concept of perception, Fuzzy Optim. Decision Making 3 (2004) 367–374.
[10] M. Kurano, M. Yasuda, J. Nakagami, Y. Yoshida, Fuzzy perceptive values for MDPs with discounting, in: V. Torra, Y. Narukawa, S. Miyamoto (Eds.), MDAI 2005, Lecture Nortes in Artificial Intelligence, Vol. 3558, Springer, Berlin, 2005, pp. 283–293.
[11] H. Mine, S. Osaki, Markov Decision Process, Elsevier, Amsterdam, 1970.
[12] E. Nummelin, General Irreducible Markov Chains and Non-negative Operators, Cambridge University Press, Cambridge, 1984.
[13] M.L. Puterman, Markov Decision Process: Discrete Stochastic Dynamic Programming, Wiley, New York, 1994.
[14] R. Sabbadin, A possibilistic model for qualitative sequential decision problems under uncertainty in partially observable environments, Uncertainty in AI Conference, 1999, pp. 567–574.
[15] R. Sabbadin, H. Fargier, J. Lang, Towards qualitative approaches to multi-stage decision making, Internat. J. Approx. Reason. 19 (1998) 441–471.
[16] D.T. Schweizer, Perturbation theory and finite Markov chains, J. Appl. Probab. 5 (1968) 401–413.
[17] E. Solan, Continuity of the value of competitive Markov decision processes, J. Theoret. Probab. 16 (2004) 831–845.
[18] Y. Yoshida, E.E. Kerre, A fuzzy ordering on multi-dimensional fuzzy sets induced from convex cones, Fuzzy Sets and Systems 130 (2002) 343–355.
[19] L.A. Zadeh, Fuzzy sets, Inform. and Control 8 (1965) 338–353.
[20] L.A. Zadeh, Toward a perception-based theory of probabilistic reasoning with imprecise probabilities, J. Statist. Plann. Inference 105 (2002) 233–264.

Please cite this article as: M. Kurano, et al., Fuzzy optimality relation for perceptive MDPs—the average case, Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss.2007.04.006