数理解析研究所講究録
945 巻 1996 年 188-198

Fuzzy Decision Processes with an Average Reward Criterion

千葉大・教育 蔵野 正美 (Masami Kurano)
千葉大・理 安田 正実 (Masami Yasuda)
千葉大・理 中神 潤一 (Jun-ichi Nakagami)
北九州大・経済 吉田 祐治 (Yuji Yoshida)

Abstract

In this paper, in the same framework as the previous papers, we will specify the long-run average fuzzy reward from any continuous fuzzy stationary policy and develop its optimization by so-called "fuzzy max order" on the bounded convex fuzzy numbers under the ergodicity and continuity conditions. By introducing the relative value functions, the average reward is characterized as a unique solution of the associated equation. Moreover, using the "vanishing discount factor" approach which is well-known in the theory of Markov decision processes we derive the optimality equation.

Keywords: Fuzzy decision processes; average reward criterion; relative value function; optimality equation.

1 Introduction and notations

In our previous paper [7], we defined Markov-type fuzzy decision processes (FDP's, for short) with a bounded fuzzy reward on the real line and developed its optimization under the discount reward criterion. Also, the long-run average fuzzy reward for some dynamic fuzzy system has been specified in our another paper [8]. However, the optimization was not given there. In this paper, in the same framework as [8] we will specify the long-run average fuzzy reward from any fuzzy policy and develop its optimization by the so-called "fuzzy max order" on the convex fuzzy numbers under the ergodicity (contraction) condition for the fuzzy state transition and the continuity condition for the fuzzy reward relation. That is, by introducing the relative value functions, the average reward from any admissible stationary policy is characterized as a unique solution of the associated equation which may be useful in the policy improvement. Moreover, using the "vanishing discount factor" approach which is well-known in the theory of Markov decision processes (for example, see [13]), we derive the optimality equation under the average fuzzy reward criterion. In the reminder of this section, we will give notations and some mathematical facts.

Let E, E_1, E_2 be convex compact subsets of some Banach space. The set of all fuzzy sets $\tilde{s}$ on E is denoted by $\mathcal{F}(E)$, which is upper semi-continuous and has a compact support with the normality condition : $\sup_{x\in E}\tilde{s}(x)=1$.

The fuzzy relation is $\tilde{p}: E_1\times E_2 \to [0,1]$ and $\tilde{p}\in\mathcal{F}(E_1\times E_2)$. The α-cut ($\alpha\in[0,1]$) of the fuzzy set $\tilde{s}$ is defined as

$$\tilde{s}_\alpha := \{x\in E \mid \tilde{s}(x)\geq\alpha\}\ (\alpha>0) \quad \text{and} \quad \tilde{s}_0 := cl\{x\in E \mid \tilde{s}(x)>0\}.$$

A fuzzy set $\tilde{s} \in \mathcal{F}(E)$ is called convex if

$$\tilde{s}(\lambda x + (1-\lambda)y) \geq \tilde{s}(x) \wedge \tilde{s}(y) \quad x, y \in E, \ \lambda \in [0,1].$$

Note that $\tilde{s}$ is convex iff the α-cut $\tilde{s}_\alpha$ is a convex set for all $\alpha \in [0,1]$ (see [4]).

A fuzzy relation $\tilde{p} \in \mathcal{F}(E_1 \times E_2)$ is called convex if

$$\tilde{p}(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \geq \tilde{p}(x_1, y_1) \wedge \tilde{p}(x_2, y_2)$$

for $x_1, x_2 \in E_1, y_1, y_2 \in E_2$, and $\lambda \in [0,1]$. The class of all convex fuzzy set is denoted by using the sub-index c as

$$\mathcal{F}_c(E) := \{\tilde{s} \in \mathcal{F}(E) \mid \tilde{s} \text{ is convex }\}.$$

The set of all non-empty closed convex subsets of E is denoted by $\mathcal{C}(E)$. Then clearly $\tilde{s} \in \mathcal{F}_c(E)$ means that $\tilde{s}_\alpha \in \mathcal{C}(E)$ for all $\alpha \in [0,1]$. Let ρ be the Hausdorff metric on $\mathcal{C}_c(E)$, then $(\mathcal{C}_c(E), \rho)$ is a complete separable metric space (c.f. [9]).

Let us restrict the term of convex fuzzy number to be convex fuzzy set with the finite support contained in the real interval $[0, M] \subset R_+ := [0, \infty)$ with a fixed positive number M, that is,

$$\mathcal{F}_c([0,M]) := \{\tilde{s} \in \mathcal{F}_c(R_+) \mid \tilde{s}_0 \subset [0,M]\},$$

and $\mathcal{C}([0,M])$ be the set of all closed convex subsets of $[0,M]$. The Hausdorff metric on $\mathcal{C}([0,M])$ is represented by δ, i.e.,

$$\delta([a,b],[c,d]) := \mid a-c \mid \vee \mid b-d \mid \quad \text{for} \quad [a,b],[c,d] \in \mathcal{C}([0,M]).$$

The addition and the multiplicative operations of fuzzy sets (fuzzy numbers) are defined as follows (see [11, 17]): For $\tilde{n}, \tilde{m} \in \mathcal{F}_c(R_+)$ and $\lambda \in R_+$, define

$$(\tilde{n} + \tilde{m})(u) := \sup_{u_1, u_2 \in R_+ :\ u_1 + u_2 = u} \{\tilde{n}(u_1) \wedge \tilde{m}(u_2)\},$$

$$(\lambda \tilde{n})(u) := \begin{cases} \tilde{n}(u/\lambda) & \text{if } \lambda > 0 \\ I_{\{0\}}(u) & \text{if } \lambda = 0, \end{cases} \quad u \in R_+.$$

It is easily seen that, for $\alpha \in (0,1]$,

$$(\tilde{n} + \tilde{m})_\alpha = \tilde{n}_\alpha + \tilde{m}_\alpha \quad \text{and} \quad (\lambda \tilde{n})_\alpha = \lambda \tilde{n}_\alpha$$

holds by this operation. Here the operation for sets means the ordinary definition as $A + B := \{x + y \mid x \in A, y \in B\}$ and $\lambda A := \{\lambda x \mid x \in A\}$ for $A, B \subset R_+$.

Lemma 1.1 ([4, Theorem 2.3]).

(i) For any $\tilde{n}, \tilde{m} \in \mathcal{F}_c(R_+)$ and $\lambda \in R_+$, $\tilde{n} + \tilde{m} \in \mathcal{F}_c(R_+)$ and $\lambda \tilde{n} \in \mathcal{F}_c(R_+)$.

(ii) For any $\tilde{s} \in \mathcal{F}_c(E_1)$ and $\tilde{p} \in \mathcal{F}_c(E_1 \times E_2)$, then $\sup_{x \in E_1} \tilde{s}(x) \wedge \tilde{p}(x, \cdot) \in \mathcal{F}_c(E_2)$.

2 Fuzzy Decision Processes

A fuzzy decision process, in this paper, is a controlled dynamic fuzzy system defined by four objects $(S, A, \tilde{q}, \tilde{r})$ as follows:

(i) Let S and A be a state space and an action space, which are given as convex compact subsets of some Banach space respectively. The decision process is assumed to be fuzzy itself, so that both the state of the system and the action taken at each stage are denoted by the element of $\mathcal{F}_c(S)$ and $\mathcal{F}_c(A)$, called the fuzzy state and the fuzzy action respectively.

(ii) The law of motion for the system and the fuzzy reward can be characterized by time invariant fuzzy relations $\tilde{q} \in \mathcal{F}_c(S \times A \times S)$ and $\tilde{r} \in \mathcal{F}_c(S \times A \times [0, M])$. Explicitly, if the system is in a fuzzy state $\tilde{s} \in \mathcal{F}_c(S)$ and the fuzzy action $\tilde{a} \in \mathcal{F}_c(A)$ is chosen, then it transfers to a new fuzzy state $Q(\tilde{s}, \tilde{a})$ and a fuzzy reward $R(\tilde{s}, \tilde{a})$ has been earned, where Q, R are defined by the following:

$$Q(\tilde{s}, \tilde{a})(y) := \sup_{(x,a)\in S\times A} \tilde{s}(x) \wedge \tilde{a}(a) \wedge \tilde{q}(x, a, y) \quad (y \in S) \tag{2.1}$$

$$R(\tilde{s}, \tilde{a})(u) := \sup_{(x,a)\in S\times A} \tilde{s}(x) \wedge \tilde{a}(a) \wedge \tilde{r}(x, a, u) \quad (0 \leq u \leq M). \tag{2.2}$$

Note that, by Lemma 1.1, it holds that $Q(\tilde{s}, \tilde{a})(\cdot) \in \mathcal{F}_c(S)$ and $R(\tilde{s}, \tilde{a})(\cdot) \in \mathcal{F}_c([0, M])$ for all $\tilde{s} \in \mathcal{F}_c(S), \tilde{a} \in \mathcal{F}_c(A)$.

Firstly we will define a policy based on the fuzzy state and fuzzy action as follows. Let $\Pi := \{\pi | \pi : \mathcal{F}_c(S) \mapsto \mathcal{F}_c(A)\}$ be the set of all maps from $\mathcal{F}_c(S)$ to $\mathcal{F}_c(A)$. Any element $\pi \in \Pi$ is called a strategy. A policy, $\check{\pi} = (\pi_1, \pi_2, \pi_3, \cdots)$, is a sequence of strategies such that $\pi_t \in \Pi$ for each t. Especially, the policy $(\pi, \pi, \pi, \cdots)$ is a stationary policy and is denoted by π^∞.

A fuzzy strategy $\pi \in \Pi$ is called admissible if the α-cut $\pi(\tilde{s})_\alpha$ of π depends only on the scalar α and the sets $\tilde{s}_\alpha$, that is, it would be written as

$$\pi(\tilde{s})_\alpha = \pi(\alpha, \tilde{s}_\alpha) \quad \text{for } \tilde{s} \in \mathcal{F}_c(S).$$

If $\pi(\alpha, D)$ is continuous in $(\alpha, D) \in [0, 1] \times \mathcal{C}(S)$, π is called continuous. We denote by Π_A and Π_C, respectively, the collections of all admissible and continuous fuzzy strategies. A policy $\check{\pi} = (\pi_1, \pi_2, \cdots)$ is called admissible (continuous resp.) if $\pi_t \in \Pi_A$ (Π_C) for all $t \geq 0$.

Definition 2.1 (c.f. [6, 11]). *For $\tilde{u}_t, \tilde{u} \in \mathcal{F}_c(E)$, $\lim_{t\to\infty} \tilde{u}_t = \tilde{u}$ iff $\lim_{t\to\infty} \sup_{\alpha\in[0,1]} \rho(\tilde{u}_{t,\alpha}, \tilde{u}_\alpha) = 0$, where $\tilde{u}_{t,\alpha}$ and $\tilde{u}_\alpha$ are respectively the α-cut of $\tilde{u}_t$ and $\tilde{u}$.*

For any closed interval $D \in \mathcal{C}([0, M])$, we put $D = [\underline{D}, \overline{D}]$, where $\underline{D}$ and $\overline{D}$ are the left and right end points of D respectively.

The partial order $\succeq$ on $\mathcal{C}([0, M])$ is defined as follows : For any $D_1, D_2 \in \mathcal{C}([0, M])$. $D_1 \succeq D_2$ means that $\overline{D_1} \geq \overline{D_2}$ and $\underline{D_1} \geq \underline{D_2}$. Then, $(\mathcal{C}([0, M]), \succeq)$ becomes a complete lattice (see [2]) and the following lemma holds obviously.

Lemma 2.1. *For any sequence $\{D_n\}_{n=1}^{\infty} \subset \mathcal{C}([0, M])$, it holds that*

(i) $\sup_{n\geq 1} D_n = [\sup_{n\geq 1} \underline{D_n}, \sup_{n\geq 1} \overline{D_n}]$, *and*

(ii) *if* $\sum_{n\geq 1}^{\infty} D_n$ *converges,* $\sum_{n\geq 1}^{\infty} D_n = [\sum_{n\geq 1}^{\infty} \underline{D_n}, \sum_{n\geq 1}^{\infty} \overline{D_n}]$.

Definition 2.2. *For any* $\tilde{n}, \tilde{m} \in \mathcal{F}_c([0, M])$, $\tilde{n} \succeq \tilde{m}$ *iff* $\tilde{n}_\alpha \succeq \tilde{m}_\alpha$ *for all* $\alpha \in [0, 1]$.

For any admissible policy $\tilde{\pi} = (\pi_1, \pi_2, \cdots)$ and an initial fuzzy state $\tilde{s} \in \mathcal{F}_c(S)$, we can define a sequence of fuzzy rewards on $[0, M]$,

$$\{R(\tilde{s}_t, \pi_t(\tilde{s}_t))\}_{t=1}^{\infty}, \quad \text{where}$$

$$\tilde{s}_1 = \tilde{s} \quad \text{and} \quad \tilde{s}_{t+1} = Q(\tilde{s}_t, \pi_t(\tilde{s}_t)) \quad \text{for } t \geq 1. \tag{2.3}$$

Here we are concerned with two performance criteria. The first one is the total discounted fuzzy reward with a discount factor β $(0 < \beta < 1)$, where the definition depends on the following lemma. The lemma is a special case of the convergence theorem in [17].

Lemma 2.2. *For any* $\tilde{s} \in \mathcal{F}_c(S)$ *and any admissible* $\tilde{\pi} = (\pi_1, \pi_2, \cdots)$,

$$\left\{\sum_{t=1}^{T} \beta^{t-1} R(\tilde{s}_t, \pi_t(\tilde{s}_t))\right\}_{T\geq 1} \tag{2.4}$$

is convergent in $\mathcal{F}_c([0, M/(1-\beta)])$.

From the above lemma, we can define the discounted total fuzzy reward as follows :

$$\psi_\beta(\tilde{\pi}, \tilde{s}) := \sum_{t=1}^{\infty} \beta^{t-1} R(\tilde{s}_t, \pi_t(\tilde{s}_t)) \in \mathcal{F}_c([0, M/(1-\beta)] \tag{2.5}$$

for $\tilde{s} \in \mathcal{F}_c(S)$ and $\tilde{\pi} = (\pi_1, \pi_2, \cdots)$.

The problem in the discounted case is to maximize $\psi_\beta(\tilde{\pi}, \tilde{s})$ over all admissible policy $\tilde{\pi}$ with respect to the order $\succeq$ on $\mathcal{F}_c([0, M])$, which has been investigated in [7]. The second performance criteria is the long-run average fuzzy reward per unit time, which is formally defined by

$$\Psi(\tilde{s}, \tilde{\pi}) := \lim_{T\to\infty} \frac{R_T}{T}, \tag{2.6}$$

$$R_T := \sum_{t=1}^{T} R(\tilde{s}_t, \pi_t(\tilde{s}_t)) \quad (T \geq 1). \tag{2.7}$$

Our problem in this paper is to show the convergency of $\Psi(\tilde{s}, \tilde{\pi})$ and maximize $\Psi(\tilde{s}, \tilde{\pi})$ over some class of continuous policies $\tilde{\pi}$ with respect to the order $\succeq$ on $\mathcal{F}_c([0, M])$, which is given in Sections 4 and 5.

3 Assumptions and preliminary results

A map $Q_\alpha : \mathcal{C}(S) \times \mathcal{C}(A) \mapsto \mathcal{C}(S)$ ($\alpha \in [0,1]$) is defined by

$$Q_\alpha(D \times B) := \begin{cases} \{y \in S \mid \tilde{q}(x,a,y) \geq \alpha \text{ for some } (x,a) \in D \times B\}, & \alpha > 0, \\ cl\{y \in S \mid \tilde{q}(x,a,y) > 0 \text{ for some } (x,a) \in D \times B\}, & \alpha = 0, \end{cases}$$

and a map $R_\alpha : \mathcal{C}(S) \times \mathcal{C}(A) \mapsto \mathcal{C}([0,M])$ ($\alpha \in [0,1]$) by

$$R_\alpha(D \times B) := \begin{cases} \{u \in R_+ \mid \tilde{r}(x,a,u) \geq \alpha \text{ for some } (x,a) \in D \times B\}, & \alpha > 0, \\ cl\{u \in R_+ \mid \tilde{r}(x,a,u) > 0 \text{ for some } (x,a) \in D \times B\}, & \alpha = 0. \end{cases}$$

Since $R_\alpha(D \times B)$ is a closed interval for each $\alpha \in [0,1]$, we can write it as

$$R_\alpha(D \times B) := [\underline{R}_\alpha(D \times B), \overline{R}_\alpha(D \times B)].$$

Assumption A (Ergodicity or contraction). *There exits γ_1 $(0 < \gamma_1 < 1)$ such that*

$$\rho(Q_\alpha(D,B), Q_\alpha(D',B)) \leq \gamma_1 \rho(D,D') \quad \textit{for all } B \in \mathcal{C}(A).$$

Assumption B (Lipschitz condition). *There exists a constant C such that*

$$|\underline{R}_\alpha(D \times B) - \underline{R}_\alpha(D' \times B)| \vee |\overline{R}_\alpha(D \times B) - \overline{R}_\alpha(D' \times B)| \leq C\rho(D,D')$$

for any $D, D' \in \mathcal{C}(S)$ and $B \in \mathcal{C}(A)$.

We now derive the optimality equation for the discounted case. Let

$$V := \{v : \mathcal{C}(S) \mapsto \mathcal{C}([0,M])\}.$$

Define a metric d_V on V by

$$d_V(v,w) := \sup_{D \in \mathcal{C}(S)} \delta(v(D), w(D)) \quad \text{for } v, w \in V.$$

Then, (V, d_V) becomes a complete metric space. Define a map $U_\alpha^\beta : V \mapsto V$ $(\alpha \geq 0)$ by

$$U_\alpha^\beta v(D) := \sup_{B \in \mathcal{C}(A)} \{R_\alpha(D \times B) + \beta v(Q_\alpha(D \times B))\} \tag{3.1}$$

for $v \in V$ and $D \in \mathcal{C}(S)$. If we write $v(D)$ and $U_\alpha^\beta v(D)$ respectively by $v(D) = [\underline{v}(D), \overline{v}(D)]$ and $U_\alpha^\beta v(D) = [\underline{U}_\alpha^\beta v(D), \overline{U}_\alpha^\beta v(D)]$, (3.1) becomes, from Lemma 2.1,

$$\underline{U}_\alpha^\beta v(D) = \sup_{B \in \mathcal{C}(A)} \{\underline{R}_\alpha(D \times B) + \beta \underline{v}(Q_\alpha(D \times B))\}, \tag{3.2}$$

$$\overline{U}_\alpha^\beta v(D) = \sup_{B \in \mathcal{C}(A)} \{\overline{R}_\alpha(D \times B) + \beta \overline{v}(Q_\alpha(D \times B))\}. \tag{3.3}$$

In [7], it is shown that the operator U_α^β is a contraction with modulus β. Thus, there exists a unique map $v_{\alpha,\beta} \in V$ such that

$$v_{\alpha,\beta} = U_\alpha^\beta v_{\alpha,\beta}. \tag{3.4}$$

Let $v_{\alpha,\beta}(D) := [\underline{v}_{\alpha,\beta}(D), \overline{v}_{\alpha,\beta}(D)]$ for all $D \in \mathcal{C}(S)$.

Lemma 3.1. *Suppose that Assumptions A and B hold. Then, we have*

$$|\underline{v}_\alpha^\beta(D) - \underline{v}_\alpha^\beta(D')| \vee |\overline{v}_\alpha^\beta(D) - \overline{v}_\alpha^\beta(D')| \le \frac{C}{1-\beta\gamma_1}\rho(D,D') \tag{3.5}$$

for all $D, D' \in \mathcal{C}(S)$.

4 Characterization of the average fuzzy reward

This section concerns the convergence of the average fuzzy reward $\Psi(\tilde{s}, \pi^\infty)$, which formally given in (2.6).

For any $\pi \in \Pi_C$, we put

$$R_T(\tilde{s}, \pi^\infty) := \sum_{t=1}^{T} R(\tilde{s}_t, \pi(\tilde{s}_t)), \tag{4.1}$$

$$\text{where} \quad \tilde{s}_1 := \tilde{s} \quad \text{and} \quad \tilde{s}_{t+1} = Q(\tilde{s}_t, \pi(\tilde{s}_t)) \ (t \ge 1).$$

Let define maps $Q_\alpha^\pi : \mathcal{C}(S) \mapsto \mathcal{C}(S)$ and $R_\alpha^\pi : \mathcal{C}(S) \mapsto \mathcal{C}([0,M])$ $(\pi \in \Pi_A, \alpha \in [0,1])$ by

$$\begin{aligned} Q_\alpha^\pi(D) &:= Q_\alpha(D \times \pi(\alpha, D)) \\ R_\alpha^\pi(D) &:= R_\alpha(D \times \pi(\alpha, D)) \end{aligned}$$

for $D \in \mathcal{C}(S)$. For $\pi \in \Pi_A$, $Q_{t,\alpha}^\pi$ $(t \ge 1)$ is defined inductively by using the composition of maps as follows :

$$\begin{aligned} Q_{1,\alpha}^\pi(D) &:= Q_\alpha^\pi(D) \\ Q_{t+1,\alpha}^\pi(D) &:= Q_{t,\alpha}^\pi Q_\alpha^\pi(D) \quad \text{for } t \ge 1 \text{ and } D \in \mathcal{C}(S). \end{aligned}$$

Lemma 4.1. *Let* $\pi \in \Pi_C$. *Then :*

(i) $\tilde{s}_{t+1,\alpha} = Q_{t,\alpha}^\pi(\tilde{s}_\alpha)$ *for* $t \ge 1$,

(ii) $R_T(\tilde{s}, \pi^\infty) \in \mathcal{F}([0, TM])$ *for* $T \ge 1$,

(iii) $(R_T(\tilde{s}, \pi^\infty))_\alpha = \sum_{t=1}^T R_\alpha(\tilde{s}_{t,\alpha}, \pi(\alpha, \tilde{s}_{t,\alpha}))$ *for* $T \ge 1$.

For any continuous strategy $\pi \in \Pi_C$, we shall say that $L(\pi)$ holds if there exist constant $\gamma\,(0 < \gamma < 1)$, $C > 0$ and a positive integer t_0 satisfying the following (i) and (ii) :

(i) $\rho(Q_{t_0,\alpha}^\pi(D), Q_{t_0,\alpha}^\pi(D')) \le \gamma\rho(D, D')$,

(ii) $\delta(R_\alpha^\pi(D), R_\alpha^\pi(D')) \le C\rho(D, D')$.

Let us denote by Π_{SC}^* the set of all $\pi \in \Pi_C$ satisfying the above condition $L(\pi)$.

Theorem 4.1. *For any $\pi \in \Pi_{SC}^*$, $\Psi(\tilde{s}, \pi^\infty)$ in (2.6) converges and satisfies the following :*

$$\Psi(\tilde{s}, \pi^\infty) = R(\tilde{p}^\pi, \pi(\tilde{p}^\pi)),$$

where $\tilde{p}^\pi \in \mathcal{F}_c(S)$ is a limiting fuzzy state satisfying that

(i) $\lim_{t\to\infty} \tilde{s}_t = \tilde{p}^\pi$, *and* $Q_\alpha^\pi(\tilde{p}_\alpha^\pi) = \tilde{p}_\alpha^\pi$ *for all* $\alpha \in [0, 1]$,

(ii) $\tilde{p}^\pi$ *is independent of the initial fuzzy state* $\tilde{s}$, *and*

(iii) $\rho(\tilde{s}_{t,\alpha}, \tilde{p}_\alpha^\pi) \le C^*\gamma^t$ *with* γ *in Assumption* $L(\pi)$ *and some* $C^* > 0$.

Theorem 4.1 says that for any $\pi \in \Pi_{SC}^*$, $\Psi(\tilde{s}, \pi^\infty)$ is independent of $\tilde{s}$, so we write it by $\Psi(\pi^\infty)$.

For simplicity, let, for each $\pi \in \Pi_{SC}^*$ and $D \in \mathcal{C}(S)$, $R_{T,\alpha}^{\pi^\infty}(D) = \sum_{t=1}^T R_\alpha^\pi(Q_{t,\alpha}^\pi(D))$. Note from Lemma 4.1 that $R_T(\tilde{s}, \pi^\infty)_\alpha = R_{T,\alpha}^{\pi^\infty}(\tilde{s}_\alpha)$ for all $T \ge 1$ and $\alpha \in [0, 1]$. Let $R_{T,\alpha}^{\pi^\infty}(D) := [\underline{R}_{T,\alpha}^{\pi^\infty}(D), \overline{R}_{T,\alpha}^{\pi^\infty}(D)]$. Then, by Lemma 2.1, we have

$$\underline{R}_{T,\alpha}^{\pi^\infty}(D) = \sum_{t=1}^T \underline{R}_\alpha^\pi(Q_{t,\alpha}^\pi(D)) \tag{4.2}$$

$$\overline{R}_{T,\alpha}^{\pi^\infty}(D) = \sum_{t=1}^T \overline{R}_\alpha^\pi(Q_{t,\alpha}^\pi(D)), \tag{4.3}$$

$$\text{where} \quad R_\alpha^\pi(D') := [\underline{R}_\alpha^\pi(D'), \overline{R}_\alpha^\pi(D')]$$

for all $D' \in \mathcal{C}(S)$. By Theorem 4.1 and Assumption B, we observe that $R_\alpha^\pi(\tilde{s}_{t,\alpha}) \to R_\alpha^\pi(\tilde{p}_\alpha^\pi)$ exponentially first as $t \to \infty$. Thus, by (4.2) and (4.3),

$$\underline{h}_\alpha^\pi(D) := \lim_{T\to\infty} (\underline{R}_{T,\alpha}^{\pi^\infty}(D) - T \times \underline{R}_\alpha^\pi(\tilde{p}_\alpha^\pi)), \tag{4.4}$$

$$\overline{h}_\alpha^\pi(D) := \lim_{T\to\infty} (\overline{R}_{T,\alpha}^{\pi^\infty}(D) - T \times \overline{R}_\alpha^\pi(\tilde{p}_\alpha^\pi)) \tag{4.5}$$

converge for all $D \in \mathcal{C}(S)$. The function $\underline{h}_\alpha$ ($\overline{h}_\alpha$ resp.) is called lower (upper) relative value function, whose basic ideas are appearing in the theory of Markov processes (c.f. [13]).

Let us denote the α-cut of the discounted fuzzy reward of (2.5):

$$\psi_\beta(\pi^\infty, \tilde{s})_\alpha = [\underline{\psi}_\beta(\pi^\infty, \tilde{s})_\alpha, \overline{\psi}_\beta(\pi^\infty, \tilde{s})_\alpha], \quad \alpha \in [0, 1]. \tag{4.6}$$

Then, for any $\pi \in \Pi_{SC}^*$, the extremal points

$$\Psi(\pi^\infty)_\alpha = [\underline{\Psi}_\alpha(\pi^\infty), \overline{\Psi}_\alpha(\pi^\infty)]$$

are characterized in the following theorem, whose description is popular in the theory of Markov decision processes (cf. [1, 13]).

Theorem 4.2. *For any* $\pi \in \Pi^*_{SC}$, *we have*

$$\underline{\psi}_\beta(\pi^\infty, \tilde{s})_\alpha = \underline{\Psi}_\alpha(\pi^\infty)/(1-\beta) + \underline{h}^\pi_\alpha(\tilde{s}_\alpha) + \underline{\varepsilon}(\beta,\alpha), \tag{4.7}$$

$$\overline{\psi}_\beta(\pi^\infty, \tilde{s})_\alpha = \overline{\Psi}_\alpha(\pi^\infty)/(1-\beta) + \overline{h}^\pi_\alpha(\tilde{s}_\alpha) + \overline{\varepsilon}(\beta,\alpha), \tag{4.8}$$

where $|\underline{\varepsilon}(\beta,\alpha)| \vee |\overline{\varepsilon}(\beta,\alpha)| \to 0$ *uniformly for* $\alpha \in [0,1]$ *as* $\beta \to 1$.

Theorem 4.3. *For any* $\pi \in \Pi^*_{SC}$, *let* $\underline{h}^\pi_\alpha$ *and* $\overline{h}^\pi_\alpha$ *be defined as (4.4) and (4.5). Then, the following equations hold :*

$$\underline{h}^\pi_\alpha(D) + \underline{\Psi}_\alpha(\pi^\infty) = \underline{R}^\pi_\alpha(D) + \underline{h}^\pi_\alpha(Q^\pi_\alpha(D)), \tag{4.9}$$

$$\overline{h}^\pi_\alpha(D) + \overline{\Psi}_\alpha(\pi^\infty) = \overline{R}^\pi_\alpha(D) + \overline{h}^\pi_\alpha(Q^\pi_\alpha(D)) \quad \text{for all } D \in \mathcal{C}(S). \tag{4.10}$$

The following theorem is useful in policy improvement.

Theorem 4.4. *For any* $\pi \in \Pi^*_{SC}$, *let* $\underline{h}^\pi_\alpha$ *and* $\overline{h}^\pi_\alpha$ *be defined as in (4.4) and (4.5). Let* $\pi' \in \Pi^*_{SC}$ *be such that*

$$\underline{h}^\pi_\alpha(D) + \underline{\Psi}_\alpha(\pi^\infty) \le \underline{R}^{\pi'}_\alpha(D) + \underline{h}^\pi_\alpha(Q^{\pi'}_\alpha(D)) \tag{4.11}$$

$$\overline{h}^\pi_\alpha(D) + \overline{\Psi}_\alpha(\pi^\infty) \le \overline{R}^{\pi'}_\alpha(D) + \overline{h}^\pi_\alpha(Q^{\pi'}_\alpha(D)) \tag{4.12}$$

for all $D \in \mathcal{C}(S)$ *and* $\alpha \in [0,1]$. *Then* $\Psi(\pi^\infty) \preceq \Psi(\pi'^\infty)$.

Corollary 4.1. *For any* $\pi \in \Pi^*_{SC}$, *let* $\underline{h}^\pi_\alpha$ *and* $\overline{h}^\pi_\alpha$ *be defined as in (4.4) and (4.5). Suppose that the following inequalities hold :*

$$\underline{h}^\pi_\alpha(D) + \underline{\Psi}_\alpha(\pi^\infty) \ge \underline{R}^{\pi'}_\alpha(D) + \underline{h}^\pi_\alpha(Q^{\pi'}_\alpha(D))$$

$$\overline{h}^\pi_\alpha(D) + \overline{\Psi}_\alpha(\pi^\infty) \ge \overline{R}^{\pi'}_\alpha(D) + \overline{h}^\pi_\alpha(Q^{\pi'}_\alpha(D))$$

for all $\pi' \in \Pi^*_{SC}$, $D \in \mathcal{C}(S)$ *and* $\alpha \in [0,1]$. *Then* π^∞ *is absolutely optimal in* Π^*_{SC}, *i.e.*,

$$\Psi(\pi^\infty) \succeq \Psi(\pi'^\infty) \quad \text{for all } \pi' \in \Pi^*_{SC}.$$

5 The optimality equation

In this section, we derive the optimality equation and consider its validity for optimization. The proof is done by the "vanishing discounted factor" method, using Arzela-Ascoli theorem (c.f. [5, 14]).

Theorem 5.1. *Suppose that Assumptions A and B in Section 3 hold. Then, for any* $\alpha \in [0,1]$, *there exist constants* $\underline{\Psi}_\alpha, \overline{\Psi}_\alpha$ *and* $\underline{v}_\alpha, \overline{v}_\alpha : \mathcal{C}(S) \mapsto \mathcal{C}([0,M])$ *such that*

$$\text{(i) } [\underline{\Psi}_{\alpha'}, \overline{\Psi}_{\alpha'}] \supset [\underline{\Psi}_\alpha, \overline{\Psi}_\alpha] \tag{5.1}$$

for any $\alpha, \alpha' \, (\alpha' < \alpha)$ *belonging to some countable subset dense in* $[0,1]$, *and*

$$\text{(ii) } \underline{v}_\alpha(D) + \underline{\Psi}_\alpha = \sup_{B \in \mathcal{C}(A)} \{\underline{R}_\alpha(D \times B) + \underline{v}_\alpha(Q_\alpha(D \times B))\} \tag{5.2}$$

$$\text{(iii) } \overline{v}_\alpha(D) + \overline{\Psi}_\alpha = \sup_{B \in \mathcal{C}(A)} \{\overline{R}_\alpha(D \times B) + \overline{v}_\alpha(Q_\alpha(D \times B))\} \tag{5.3}$$

for all $D \in \mathcal{C}(S)$ *and* $\alpha \in [0,1]$.

It will be shown in Theorem 5.2 that $[\underline{\Psi}_\alpha, \overline{\Psi}_\alpha]$ given in Theorem 5.1 is corresponding to the α-cut of the maximum average fuzzy reward, so that (5.2) and (5.3) are interpreted as the optimality equations for our average fuzzy decision model. For this purpose, we need the following lemmas.

Lemma 5.1. *For any* $\tilde{n}, \tilde{m} \in \mathcal{F}_c([0,M])$, *if* $\tilde{n}_\alpha \succeq (=) \tilde{m}_\alpha$ *on some subset* F *dense in* $[0,1]$, *then* $\tilde{n} \succeq (=) \tilde{m}$.

Lemma 5.2 (c.f. [6, 12]). *Suppose that a family of subsets* $\{D_\alpha, \alpha \in [0,1]\}$ *satisfies the following* (i) *and* (ii) *:*

(i) $D_{\alpha'} \subset D_\alpha$ *for all* $\alpha, \alpha' (0 \le \alpha' \le \alpha \le 1)$,

(ii) $\lim_{\alpha' \uparrow \alpha} D_{\alpha'} = D_\alpha$ *for all* $\alpha \in [0,1]$.

Then, $\tilde{s}(x) := \sup_{\alpha \in [0,1]} \{\alpha \wedge I_{D_\alpha}(x)\}$, $x \in S$, *satisfies* $\tilde{s} \in \mathcal{F}(S)$ *and* $\tilde{s}_\alpha = D_\alpha$ *for all* $\alpha \in [0,1]$.

Let denote by F a countable subset on which (i) in Theorem 5.1 holds. For $\underline{\Psi}_\alpha$ and $\overline{\Psi}_\alpha$ given in Theorem 5.1, let, for any $\alpha \in [0,1]$,

$$\Psi_\alpha^- := \lim_{\alpha' \uparrow \alpha \text{ with } \alpha' \in F} \underline{\Psi}_{\alpha'}, \quad \Psi_0^- := \underline{\Psi}_0,$$

$$\Psi_\alpha^+ := \lim_{\alpha' \uparrow \alpha \text{ with } \alpha' \in F} \overline{\Psi}_{\alpha'}, \quad \Psi_0^+ := \overline{\Psi}_0.$$

Since the conditions (i) and (ii) in Lemma 5.2 hold for the family $\{[\Psi_\alpha^-, \Psi_\alpha^+], \alpha \in [0,1]\}$, we can construct the fuzzy set $\tilde{\Psi}$ by

$$\tilde{\Psi}(x) := \sup_{\alpha \in [0,1]} \{\alpha \wedge 1_{[\Psi_\alpha^-, \Psi_\alpha^+]}(x)\} \quad x \in S.$$

Theorem 5.2.

(i) $\tilde{\Psi} \succeq \Psi(\pi^\infty)$ *for any* $\pi \in \Pi^*_{SC}$.

(ii) *If there exists a strategy* $\pi^* \in \Pi^*_{SC}$ *such that*

$$\underline{v}_\alpha(D) + \underline{\Psi}_\alpha = \underline{R}_\alpha(D \times \pi^*(\alpha, D)) + \underline{v}_\alpha(Q_\alpha(D \times \pi^*(\alpha, D)))$$

$$\overline{v}_\alpha(D) + \overline{\Psi}_\alpha = \overline{R}_\alpha(D \times \pi^*(\alpha, D)) + \overline{v}_\alpha(Q_\alpha(D \times \pi^*(\alpha, D)))$$

for all $D \in \mathcal{C}(S)$ *and* $\alpha \in [0,1]$, *then* $\pi^{*\infty}$ *is absolutely optimal in* Π^*_{SC}, *i.e.,*

$$\Psi(\pi^{*\infty}) \succeq \Psi(\pi^\infty) \quad \text{for all } \pi \in \Pi^*_{SC}.$$

References

[1] D.P.Bertsekas and S.E.Shreve, *Stochastic Optimal Control : the Discrete Time Case* (Academic Press, New York, 1978).

[2] G.Birkhoff, Lattice Theory, *Amer. Math. Soc., Coll. Pub.*, **25** (1940).

[3] D.Blackwell, Discrete dynamic programming, *AMS* **33** (1962) 719-726.

[4] Chen-wei Xu, On convexity of fuzzy sets and fuzzy relations, *Information Science* **59** (1992) 92-102.

[5] E.Fernándes-Gaucherand, A note on Ross-Taylor theorem, *Applied Math. and Computation* **64** (1994) 207-212.

[6] M.Kurano, M.Yasuda, J.Nakagami and Y.Yoshida, A limit theorem in some dynamic fuzzy systems, *Fuzzy Sets and Systems* **51** (1992) 83-88.

[7] M.Kurano, M.Yasuda, J.Nakagami and Y.Yoshida, Markov-type fuzzy decision processes with a discounted reward on a closed interval, *Technical Reports of Mathematical Sciences, Chiba University,* No.10, Vol.4 (1992).

[8] M.Kurano, M.Yasuda, J.Nakagami and Y.Yoshida, The time average reward with some dynamic fuzzy systems, *Technical Reports of Mathematical Sciences, Chiba University,* No.10, Vol.5 (1992).

[9] K.Kuratowski, *Topology I* (Academic Press, New York, 1966).

[10] M.L.Puri and D.A.Ralescu, The concept of normality for fuzzy random variables, *Ann. Prob.* **13** (1985) 1373-1379.

[11] S.Nanda, On sequences of fuzzy numbers, *Fuzzy Sets and Systems* **33** (1989) 123-126.

[12] V.Novàk, *Fuzzy Sets and Their Applications* (Adam Hilder, Bristol-Boston, 1989).

[13] S.M.Ross, *Applied probability models with optimization applications* (Holden-Say, San Francisco, 1970).

[14] H.L.Royden, *Real Analysis*, 2nd ed. (Macmillan, New York, 1968).

[15] H.M.Taylor, Markovian sequential replacement processes, *Ann. Math. Statist.* **38** (1965) 1677-2122.

[16] M.Yasuda, A linear structure and convexity for relations in dynamic fuzzy systems, *Computers Math. Applic.* **27** (1994) 23-27.

[17] Y.Yoshida, M.Yasuda, J.Nakagami and M.Kurano, A potential of fuzzy relations with a linear structure : The contractive case, *Fuzzy Sets and Systems* **60** (1993) 283-294.

[18] Y.Yoshida, M.Yasuda, J.Nakagami and M.Kurano, A potential of fuzzy relations with a linear structure : The unbounded case, *Fuzzy Sets and Systems* **66** (1994) 83-95.

[19] Y.Yoshida, The recurrence of dynamic fuzzy systems, *RIFIS Technical Report*, No. 88 (1994), Research Institute of Fundamental Information Science Kyushu University 33 Fukuoka 812, Japan.

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