数理解析研究所講究録
第 864 巻 1994 年 180-189

Markov–Type Fuzzy Decision Processes with a Discounted Reward on a Closed Interval

千葉大・教育　蔵野 正美 (Masami Kurano)
千葉大・教養　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

We will define a new multi-stage decision process, which is termed Markov-type fuzzy decision process. By the general framework in the decision process, the optimization problem of the discounted reward is discussed under a partial order of convex fuzzy numbers.

Keywords: Fuzzy decision, Markov-type, partial order, policy improvement, convex fuzzy relation, contractive case, uniform continuity

1 Introduction

Fuzzy decision making originated by Bellman and Zadeh[2] is a multi-stage process in which the goals and/or the constraints are fuzzy. The method of the dynamic programming is shown to be a powerful computational technique for these problem[2,7]. There are many papers on the analysis of static arguments for the fuzzy theory, however there are a few on optimization of fuzzy dynamic system. It is desirable that the fuzzification of the state of the system and its transition by fuzzy relation are are well defined, and a multi-stage decision processes is developed for a wide application.

In this paper, by trying to do fuzzification of the system, we will define a new multi-stage decision process with Markov-type fuzzy transition. (see [9,14] for Markov-type fuzzy transition). The optimization of the discount total reward for the processes under some partial order, called "fuzzy max order", on the class of convex fuzzy numbers is considered.

The analysis is done by operators on some class of functions, which is popular in Markov decision processes(for example, see[4,5]) and, applying Banach's fixed point theorem, the discounted total fuzzy reward incurred by any fuzzy policy satisfying some reasonable conditions is obtained as a unique solution of related fuzzy relational equations. Also, optimality fuzzy relational equation is given to characteristic optimal fuzzy policy.

In section 2, we list the notations and construct the model to be analyzed for the purpose. In sections 3, the functional characterization of the discounted total fuzzy reward is given and several results useful in policy improvement are obtained. The optimization is done in section 4, in which the fuzzy optimality equation is studied under some continuity conditions.

2 Notations and Assumptions

In this section, we shall give notations and mathematical facts in order to formulate a fuzzy decision processes considered in the sequel. Let E, E_1, E_2 be convex compact subsets of some Banach space. Throughout the paper we will denote a fuzzy set and a fuzzy relation by their membership functions. Refer to Zadeh[14] and Novák[12] for the theory of fuzzy sets.

The set of all fuzzy sets $\tilde{s}$ on E is denoted by $\mathcal{F}(E)$, which are assumed that it is upper semi-continuous and have a compact support with the normality condition: $\sup_{x\in E}\tilde{s}(x) = 1$ throughout the paper. The fuzzy relation between the space E_1 and E_2 means that $\tilde{p} : E_1\times E_2 \to [0,1]$ and $\tilde{p} \in \mathcal{F}(E_1 \times E_2)$. Its α-cut ($\alpha \in [0,1]$) of the fuzzy set $\tilde{s}$, $\tilde{s}_\alpha$, is defined as

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

where cl denotes the closure on the set. A fuzzy set $\tilde{s} \in \mathcal{F}(E)$ is called convex if

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

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]). Some papers un convex analysis call this notion as quasi-convex.

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 convex closed subset 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 $\alpha \in [0,1]$.

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

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

Let $\mathcal{C}([0, M])$ be a set of all closed convex subsets of $[0, M]$. For a non-empty closed interval, the Hausedroff metric δ could be considered and it becomes a complete separable metric space, 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 operation of fuzzy sets(fuzzy numbers) are defined as follows(see [7]): For $\tilde{n}, \tilde{m} \in \mathcal{F}_c(R_+)$ and $\lambda \in R_+$, define

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

and

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

where $\lambda \wedge \mu := \min\{\lambda, \mu\}$ for scalars λ, μ and $I_A(\cdot)$ means the classical indicator function of a set $A \subset R_+$. 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 $A, B \subset R_+$.

The following results have appeared in Chen-wei Xu[6].

Lemma 2.1 ([5,Th.2.3],[6]).

(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 $\sup_{x \in E_1} \tilde{s}(x) \wedge \tilde{p}(x, \cdot) \in \mathcal{F}_c(E_2)$.

Now, we consider Markov-type fuzzy decision processes

$$(S, A, [0, M], \tilde{q}, \tilde{r}, \beta)$$

which satisfy the following (i) — (iii):

(i) Let S and A be a state space and a 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 $\tilde{r} \in \mathcal{F}_c(S \times A \times [0, M])$, where M is a given positive number. 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 incurred, where Q, R are defined by the following:

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

and

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

Note that, by Lemma 2.1, 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 $\tilde{s} \in \mathcal{F}_c(S), \tilde{a} \in \mathcal{F}_c(A)$.

(iii) The constant scaler β is a discount rate satisfying $0 < \beta < 1$.

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 a 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, $\breve{\pi} = (\pi_1, \pi_2, \pi_3, \cdots)$, 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 denoted by π^∞.

For any policy $\breve{\pi} = (\pi_1, \pi_2, \cdots)$ and any initial fuzzy state $\tilde{s} \in \mathcal{F}_c(S)$, we define sequentially the fuzzy state $\{\tilde{s}_t\}$ as

$$\tilde{s}_1 := \tilde{s}, \quad \tilde{s}_{t+1} := Q(\tilde{s}_t, \pi_t(\tilde{s}_t)) \quad \text{for} \quad t = 1, 2, \cdots \tag{2.3}$$

The transition of fuzzy states by (2.3) has Markov property, that is, the state of $t + 1^{th}$ step is determined by that of t^{th} step, so that the decision processes defined above could be called Markov-type, as is said in the title.

To describe the discounted total fuzzy reward incurred from a fuzzy policy $\breve{\pi}$, let us consider the convergence of a sequence of fuzzy numbers belonging to $\mathcal{F}(R_+)$.

Definition 1.1 ([9, 11]). For $\tilde{n}_t, \tilde{n} \in \mathcal{F}_c(R_+)$,

$$\lim_{t\to\infty} \tilde{n}_t = \tilde{n}$$

iff $\lim_{t\to\infty} \sup_{\alpha\in[0,1]} \delta(\tilde{n}_{t,\alpha}, \tilde{n}_\alpha) = 0$.

The following lemma is a special case of convergence theorem proved in [13].

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

$$\left\{ \sum_{t=1}^{T} \beta^{t-1} R(\tilde{s}_t, \pi_t(\tilde{s}_t)) \right\}_{T\ge1} \tag{2.4}$$

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

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

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

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

The problem is to maximize the fuzzy reward $\psi(\breve{\pi}, \tilde{s})$ over a certain class of fuzzy policies $\{\breve{\pi}\}$ with respect to a given partial order on $\mathcal{F}_c([0, M/(1-\beta)])$. In the sequel, the problem is analyzed by introducing the partial order called "fuzzy max order".

Remarks: The fuzzy decision process defined in the previous argument is compatible with the extension principle of Zadeh[14], which gives a natural extension of non-fuzzy systems. To explain the notion, we treat the following usual deterministic systems : For given $x_t \in S$, $a_t \in A$, the next describes motion of the system;

$$x_{t+1} = f(x_t, a_t) \qquad (t = 1, 2, \cdots), \tag{2.5}$$

where $x_1 \in S$ is an initial state and $f : S \times A \mapsto S$ is a deterministic transition function. A fuzzy relation $\tilde{f} \in \mathcal{F}(S \times A \times S)$ will be defined by

$$\tilde{f}(x, a, y) = \begin{cases} 1 & \text{if } y = f(x, a) \\ 0 & \text{if } y \neq f(x, a) \end{cases}$$

Then, using the above $\tilde{f}$, the deterministic system can be rewritten. For $\tilde{x}_t \in \mathcal{F}(S)$, $\tilde{a}_t \in \mathcal{F}(A)$,

$$\tilde{x}_{t+1} = Q(\tilde{x}_t, \tilde{a}_t) \qquad (t = 1, 2, \cdots),$$

$\tilde{x}_1 = 1_{\{x_1\}}$, $\tilde{a}_1 = 1_{\{a_1\}}$ and

$$Q(\tilde{x}_t, \tilde{a}_t)(y) = \sup_{(x,a) \in S \times A} \tilde{x}_t(x) \wedge \tilde{a}_t(a) \wedge \tilde{f}(x, a, y) \quad (y \in S).$$

This shows that the transition (2.1) of fuzzy states is a fuzzy extension of the deterministic system (2.5) in extending the state space S and action space A to fuzzy sets $\mathcal{F}(S)$ and $\mathcal{F}(A)$, respectively.

3 Partial order for the optimization

We introduce a partial order on $\mathcal{F}_c([0, M])$ and give some results on the optimization of the fuzzy decision processes defined in the previous section. For $\tilde{n}, \tilde{m} \in \mathcal{F}_c([0, M])$, a partial order for fuzzy numbers is defined as

$$\tilde{n} \succeq \tilde{m}$$

if $\min \tilde{n}_\alpha \geq \min \tilde{m}_\alpha$ and $\max \tilde{n}_\alpha \geq \max \tilde{m}_\alpha$ for all $\alpha \in [0, 1]$ where min, max means the left or right end point of the α-cut intervals respectively. It is immediate that $(\mathcal{F}_c([0, M]), \succeq)$ becomes a complete lattice (see [3],[9]). Note also that

$$\sup_t \tilde{u}_t \in \mathcal{F}_c([0, M]) \quad \text{for } \{\tilde{u}_t\} \subset \mathcal{F}_c([0, M])$$

holds, where the supremum is taken with respect to the order $\succeq$.

Definition 3.1. The fuzzy strategy $\pi : \mathcal{F}_c(S) \mapsto \mathcal{F}_c(A)$ is called admissible if the α-cut set $\pi(\tilde{s})_\alpha$ of π depends only on the scaler α and the set $\tilde{s}_\alpha$, that is, it could be written as

$$\pi(\tilde{s})_\alpha = \pi(\alpha, \tilde{s}_\alpha). \tag{3.1}$$

Let Π_A be the collection of all admissible fuzzy strategies. Similarly a policy $\check{\pi} = (\pi_1, \pi_2, \cdots)$ is called admissible if $\pi_t \in \Pi_A$ $(t = 1, 2, \cdots)$.

Assumption 1. It is sufficient for the optimization of $\psi(\check{\pi}, \tilde{s})$ to deal with admissible policies.

Our problem is to maximize $\psi(\check{\pi}, \tilde{s})$ over all admissible policies $\check{\pi}$ with respect to the order $\succeq$ on $\mathcal{F}_c([0, M])$.

In order to discuss the fuzzy transition and the fuzzy reward, some notations are introduced. A map $\tilde{q}_\alpha : \mathcal{C}(S) \times \mathcal{C}(A) \mapsto \mathcal{C}(S)$ $(\alpha \in [0, 1])$ is defined by

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

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

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

By using $\tilde{q}$ and $\tilde{r}$, 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

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

for $D \in \mathcal{C}(S)$. For any admissible fuzzy policy $\tilde{\pi} = (\pi_1, \pi_2, \cdots)$, $Q_{t,\alpha}^{\tilde{\pi}}$ $(t \geq 1)$ is defined inductively by using the composition of maps as follows:

$$Q_{0,\alpha}^{\tilde{\pi}}(D) := I \ (identity), \quad Q_{1,\alpha}^{\tilde{\pi}}(D) := Q_\alpha^{\pi_1}(D)$$

and

$$Q_{t+1,\alpha}^{\tilde{\pi}}(D) := Q_\alpha^{\pi_1} Q_{t,\alpha}^{(\pi_2,\pi_3,\cdots)}(D) := Q_{t,\alpha}^{(\pi_2,\pi_3,\cdots)} Q_\alpha^{\pi_1}(D)$$

for $t = 1, 2, \cdots$ and $D \in \mathcal{C}(S)$.

Then, the following lemma holds regarding α-cuts of fuzzy states and fuzzy reward for each step.

Lemma 3.1. Let $\tilde{s} \in \mathcal{F}_c(S)$ and $\tilde{\pi} = (\pi_1, \pi_2, \cdots)$ be any admissible policy. Then, for $t = 1, 2, \cdots$ and $\alpha \in [0, 1]$,

(i)

$$\tilde{s}_{t+1,\alpha} = Q_{t,\alpha}^{\tilde{\pi}}(\tilde{s}_\alpha);$$

(ii)

$$R(\tilde{s}_t, \pi_t(\tilde{s}_t))_\alpha = R_\alpha^{\pi_t}(\tilde{s}_{t,\alpha});$$

(iii)

$$\psi(\tilde{\pi}, \tilde{s})_\alpha = \sum_{t=0}^{\infty} \beta^t R(\tilde{s}_t, \pi_t(\tilde{s}_t))_\alpha.$$

Proof. (i) For $t = 1$, let $\pi = \pi_1 \in \Pi_A$ and $\alpha \in [0, 1]$. we have

$$Q_{1,\alpha}^{\tilde{\pi}}(\tilde{s}_\alpha) = Q_\alpha^\pi(\tilde{s}_\alpha) = \tilde{q}_\alpha(\tilde{s}_\alpha \times \pi(\alpha, \tilde{s}_\alpha)) = \tilde{q}_\alpha(\tilde{s}_\alpha \times \pi(\tilde{s})_\alpha) = \tilde{q}(\tilde{s}, \pi(\tilde{s}))_\alpha = \tilde{s}_{2,\alpha}.$$

Inductively we obtain the result for $t = 2, 3, \cdots$. (ii) For $t = 1, 2, \cdots$ we have $R_\alpha^{\pi_t}(\tilde{s}_{t,\alpha}) = \tilde{r}_\alpha(\tilde{s}_{t,\alpha} \times \pi_t(\alpha, \tilde{s}_{t,\alpha})) = \tilde{r}_\alpha(\tilde{s}_{t,\alpha} \times \pi_t(\tilde{s}_t)_\alpha) = R(\tilde{s}_t, \pi_t(\tilde{s}_t))_\alpha$. by letting $D = \tilde{s}_{t,\alpha}$. (iii) It is immediately obtained from (ii) and the convexity of the fuzzy set. *q.e.d.*

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) is a complete metric space. For $v, w \in V$, we define an order

$$v \succeq w$$

by $v(D) \succeq_{\mathrm{ci}} w(D)$ for all $D \in \mathcal{C}(S)$, where $\succeq_{\mathrm{ci}}$ means that $[a, b] \succeq_{\mathrm{ci}} [c, d]$ iff $a \geq c$ and $b \geq d$ for closed intervals $\mathcal{C}([0, M])$. Further define a map $U_\alpha^\pi : V \mapsto V$ $(\pi \in \Pi_A, \alpha \in [0,1])$ by

$$U_\alpha^\pi v(D) := R_\alpha^\pi(D) + \beta v\left(Q_\alpha^\pi(D)\right) \tag{3.2}$$

for $v \in V$ and $D \in \mathcal{C}(S)$.

We will prove the contractive property of the operator U_α^π. To do it, it needs the following lemma whose proof is easy.

Lemma 3.2.

(i) Let $\Gamma := \{\gamma\}$ be an index set and $[a_\gamma, b_\gamma], [c_\gamma, d_\gamma] \in \mathcal{C}([0, M/(1-\beta)])$ for $\gamma \in \Gamma$. Then

$$\delta(\sup_{\gamma\in\Gamma}[a_\gamma, b_\gamma], \sup_{\gamma\in\Gamma}[c_\gamma, d_\gamma]) \leq \sup_{\gamma\in\Gamma} \delta([a_\gamma, b_\gamma], [c_\gamma, d_\gamma]).$$

(ii) If $[a_1, b_1], [c_1, d_1], [a_2, b_2], [c_2, d_2] \in \mathcal{C}([0, M])$, then

$$\begin{aligned} & \delta([a_1, b_1] + [c_1, d_1], [a_2, b_2] + [c_2, d_2]) \\ \leq \;& \delta([a_1, b_1], [a_2, b_2]) + \delta([c_1, d_1], [c_2, d_2]). \end{aligned}$$

(iii) If $[a, b], [c, d] \in \mathcal{C}([0, M])$, then

$$\delta(\beta[a, b], \beta[c, d]) = \beta\delta([a, b], [c, d]).$$

Now can state our main results in this section.

Theorem 3.1. Let $\pi \in \Pi_A$ and $\alpha \in [0, 1]$. It holds that U^π_α is monotone, contractive and has a unique map $v^\pi_\alpha \in V$ such that

$$v^\pi_\alpha = U^\pi_\alpha v^\pi_\alpha. \tag{3.3}$$

Proof. Fix any $\pi \in \Pi_A$ and $\alpha \in [0, 1]$. For any $v, w \in V$, since

$$\begin{aligned} U^\pi_\alpha v(D) &= R^\pi_\alpha(D) + \beta v(Q^\pi_\alpha(D)), \\ U^\pi_\alpha w(D) &= R^\pi_\alpha(D) + \beta w(Q^\pi_\alpha(D)), \end{aligned}$$

and Lemma 3.2, we have

$$\begin{aligned} \delta(U^\pi_\alpha v(D), U^\pi_\alpha w(D)) &\leq \delta(R^\pi_\alpha(D)), R^\pi_\alpha(D)) + \delta(\beta v(Q^\pi_\alpha(D)), \beta w(Q^\pi_\alpha(D))) \\ &= \beta\delta(v(Q^\pi_\alpha(D)), w(Q^\pi_\alpha(D))) \\ &\leq \beta d_V(v, w) \end{aligned}$$

for all $D \in \mathcal{C}(S)$. This means

$$d_V(U^\pi_\alpha v, U^\pi_\alpha w) \leq \beta d_V(v, w).$$

That is, U^π_α is contractive and, by Banach's fixed point theorem, has a unique map $v^\pi_\alpha \in V$ such that $v^\pi_\alpha = U^\pi_\alpha v^\pi_\alpha$. Further if $v \succeq_V w$, then we have

$$\begin{aligned} U^\pi_\alpha v(D) &= R^\pi_\alpha(D) + \beta v(Q^\pi_\alpha(D)) \\ &\succeq_{\text{ci}} R^\pi_\alpha(D) + \beta w(Q^\pi_\alpha(D)) \\ &= U^\pi_\alpha w(D) \end{aligned}$$

for all $D \in \mathcal{C}(S)$. So $U^\pi_\alpha v \succeq_V U^\pi_\alpha w$. Therefore U^π_α is monotone. *q.e.d.*

Theorem 3.2. For $\tilde{s} \in \mathcal{F}_c(S)$ and a admissible stationary policy $\pi^\infty = (\pi, \pi, \pi, \cdots)$,

$$\psi(\pi^\infty, \tilde{s})_\alpha = v^\pi_\alpha(\tilde{s}_\alpha)$$

holds for $\alpha \in [0, 1]$.

Proof. Let $\tilde{s} \in \mathcal{F}_c(S)$, $\pi^\infty = (\pi, \pi, \pi, \cdots)$ and $\alpha \in [0, 1]$. And define ψ_α by

$$\psi_\alpha(\pi^\infty, D) := \sum_{t=0}^{\infty} \beta^t R^\pi_\alpha(Q^{\pi^\infty}_{t,\alpha}(D))$$

for $D \in \mathcal{C}(S)$. Then from Lemma 3.1(ii) we have

$$\begin{aligned} \psi(\pi^\infty, \tilde{s})_\alpha &= \sum_{t=0}^{\infty} \beta^t R(\tilde{s}_t, \pi(\tilde{s}_t))_\alpha \\ &= \sum_{t=0}^{\infty} \beta^t R^\pi_\alpha(\tilde{s}_{t,\alpha}) \\ &= \psi_\alpha(\pi^\infty, \tilde{s}_\alpha). \end{aligned}$$

From Lemma 3.1(i),(ii) we have

$$\begin{aligned}
\psi_\alpha(\pi^\infty, \tilde{s}_\alpha) &= \sum_{t=0}^{\infty} \beta^t R_\alpha^\pi(\tilde{s}_{t,\alpha}) \\
&= R_\alpha^\pi(\tilde{s}_\alpha) + \beta\{\sum_{t=1}^{\infty} \beta^{t-1} R_\alpha^\pi(Q_{t-1,\alpha}^\pi(\tilde{s}_{t-1,\alpha}))\} \\
&= R_\alpha^\pi(\tilde{s}_\alpha) + \beta\psi_\alpha(\pi^\infty, Q_\alpha^\pi(\tilde{s}_\alpha)) \\
&= U_\alpha^\pi \psi_\alpha(\pi^\infty, \tilde{s}_\alpha).
\end{aligned}$$

Therefore $\psi_\alpha(\pi^\infty, \tilde{s}_\alpha) = U_\alpha^\pi \psi_\alpha(\pi^\infty, \tilde{s}_\alpha)$. From Theorem 3.1 and (3.3), we obtain $\psi(\pi^\infty, \tilde{s})_\alpha = \psi_\alpha(\pi^\infty, \tilde{s}_\alpha) = v_\alpha^\pi(\tilde{s}_\alpha)$. *q.e.d.*

Lemma 3.3. Let $\tilde{s} \in \mathcal{F}_c(S)$ be any initial fuzzy state and let $\check{\pi} = (\pi_1, \pi_2, \cdots)$ be any policy. Then

(i)
$$\psi_\alpha(\check{\pi}, \tilde{s}_\alpha) = U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} \psi_\alpha(\check{\pi}_t, \tilde{s}_\alpha)$$
where $\check{\pi}_t = (\pi_t, \pi_{t+1}, \pi_{t+2}, \cdots)$.

(ii) For any $v \in V$,
$$\psi_\alpha(\check{\pi}, \tilde{s}_\alpha) = \lim_{t\to\infty} U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} v(\tilde{s}_\alpha).$$

Proof. (i) It is similar to the proof of Theorem 3.2. (ii) Similar to Theorem 3.1 we have

$$\begin{aligned}
&\delta(U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} \psi_\alpha(\check{\pi}_t, \tilde{s}_\alpha),\ U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} v(\tilde{s}_\alpha)) \\
\le\ &\beta^{t-1} \delta(\psi_\alpha(\check{\pi}_t, \tilde{s}_\alpha),\ v(\tilde{s}_\alpha)).
\end{aligned}$$

Therefore we obtain the result. *q.e.d.*

Theorem 3.3. Let $\check{\pi} = (\pi_1, \pi_2, \cdots)$ be any policy. Suppose

$$\psi_\alpha(\check{\pi}, D) \succeq_{\text{ci}} U_\alpha^\pi \psi_\alpha(\check{\pi}, D) \quad \text{for all} \quad D \in \mathcal{C}(S), \pi \in \Pi_A \text{ and } \alpha \in [0,1]. \tag{3.4}$$

Then we have
$$\psi(\check{\pi}, \tilde{s}) \succeq \psi(\check{\sigma}, \tilde{s}) \quad \text{for all} \quad \tilde{s} \in \mathcal{F}_c(S) \text{ and any policy } \check{\sigma}.$$

Proof. By Lemma 3.3 and the monotonicity of U_α^π, it could be shown easily that $\psi_\alpha(\check{\pi}, D) \succeq_{\text{ci}} \psi_\alpha(\check{\sigma}, D)$ for all $D \in \mathcal{C}(S)$, $\alpha \in [0,1]$ and any policy $\check{\sigma}$. Therefore we have obtained the theorem. *q.e.d.*

Theorem 3.4. Let $\check{\pi} = (\pi_1, \pi_2, \cdots)$ be any policy and let $\pi \in \Pi_A$. Suppose

$$U_\alpha^\pi \psi_\alpha(\check{\pi}, D) \succeq_{\text{ci}} \psi_\alpha(\check{\pi}, D) \quad \text{for all } D \in \mathcal{C}(S) \text{ and } \alpha \in [0,1].$$

Then we have
$$\psi(\pi^\infty, \tilde{s}) \succeq \psi(\check{\pi}, \tilde{s}) \quad \text{for all } \tilde{s} \in \mathcal{F}_c(S).$$

Proof. Similarly to Theorem 3.3, it could be shown easily. *q.e.d.*

Remark. The results like Theorem 3.3 and 3.4 have appeared already in the classic discounted Markov decision model and used for the policy improvement([4,5]). By the same idea, the above theorems would be useful for the policy improvement under the fuzzy decision model.

4 Optimality equation

The objective in this section is to give a fuzzy optimality equation which is used in the optimization of the decision processes.

Define a map $U_\alpha : V \mapsto V$ $(\alpha > 0)$ by

$$U_\alpha v(D) := \sup_{B\in\mathcal{C}(A)} \{\tilde{r}_\alpha(D\times B) + \beta v(\tilde{q}_\alpha(D\times B))\} \tag{4.1}$$

for $v \in V$ and $D \in \mathcal{C}(S)$, where sup is taken over the order $\succeq$.

Theorem 4.1. Let $\alpha \in [0,1]$. U_α is monotone, contractive and has a unique map $v_\alpha^* \in V$ such that

$$v_\alpha^* = U_\alpha v_\alpha^*. \tag{4.2}$$

Proof. Using Lemma 2.2, for $v, w \in V$ we obtain

$$\begin{aligned}\delta(U_\alpha v(D), U_\alpha w(D)) &\le \sup_{B\in\mathcal{C}(A)} \delta(\beta v(\tilde{q}_\alpha(D\times B), \beta w(\tilde{q}_\alpha(D\times B)) \\ &= \beta \sup_{B\in\mathcal{C}(A)} \delta(v(\tilde{q}_\alpha(D\times B), w(\tilde{q}_\alpha(D\times B)) \\ &\le \beta d_V(v,w)\end{aligned}$$

for $D \in \mathcal{C}(S)$. Therefore $d_V(U_\alpha v, U_\alpha w) \le \beta d_V(v,w)$. By contraction, there exists a unique $v_\alpha^* \in V$ such that $v_\alpha^* = U_\alpha v_\alpha^*$. *q.e.d.*

Put, for $\alpha \in [0,1]$,

$$\begin{aligned}E_\alpha^{\tilde{r}} &:= \{(x,a,u) \in S\times A\times[0, M/(1-\beta)] \mid \tilde{r}(x,a,u) \ge \alpha\}, \\ E_\alpha^{\tilde{q}} &:= \{(x,a,y) \in S\times A\times S \mid \tilde{q}(x,a,y) \ge \alpha\}.\end{aligned}$$

For $\epsilon > 0$ we define their ϵ-covering of each set by

$$\begin{aligned}E_\alpha^{\tilde{r}}(\epsilon) &:= \{(x',a',u') \in S\times A\times[0, M/(1-\beta)] \mid \|(x',a',u') - (x,a,u)\| < \epsilon \\ &\qquad \text{for some } (x,a,u) \in E_\alpha^{\tilde{r}}\}, \\ E_\alpha^{\tilde{q}}(\epsilon) &:= \{(x',a',y') \in S\times A\times S \mid \|(x',a',y') - (x,a,y)\| < \epsilon \\ &\qquad \text{for some } (x,a,y) \in E_\alpha^{\tilde{q}}\}.\end{aligned}$$

Assumption 2 (A uniform continuity on $\tilde{r}$ and $\tilde{q}$). There exists $\eta : [0,\infty) \mapsto [0,\infty)$ such that

(i) $\eta(t) \to 0$ as $t \downarrow 0$;

(ii) $E_\alpha^{\tilde{r}}(\eta(|\alpha'-\alpha|)) \supset E_{\alpha'}^{\tilde{r}}$ for $0 \le \alpha' < \alpha$;

(iii) $E_\alpha^{\tilde{q}}(\eta(|\alpha'-\alpha|)) \supset E_{\alpha'}^{\tilde{q}}$ for $0 \le \alpha' < \alpha$.

Lemma 4.1. Suppose Assumption 2. Let $D_{\alpha'} \downarrow D_\alpha$ as $\alpha' \uparrow \alpha$ for $D_{\alpha'} \in \mathcal{C}(S)$ and $\alpha \in (0,1]$. Then

(i) $\sup_{B\in\mathcal{C}(A)} \delta(\tilde{r}_{\alpha'}(D_{\alpha'}\times B), \tilde{r}_\alpha(D_\alpha\times B)) \to 0$ as $\alpha' \uparrow \alpha$;

(ii) $\sup_{B\in\mathcal{C}(A)} \rho(\tilde{q}_{\alpha'}(D_{\alpha'}\times B), \tilde{q}_\alpha(D_\alpha\times B)) \to 0$ as $\alpha' \uparrow \alpha$.

Proof. (i) We have

$$\begin{aligned}&\delta(\tilde{r}_{\alpha'}(D_{\alpha'}\times B), \tilde{r}_\alpha(D_\alpha\times B)) \\ \le\ & \delta(\tilde{r}_{\alpha'}(D_{\alpha'}\times B), \tilde{r}_\alpha(D_{\alpha'}\times B)) + \delta(\tilde{r}_\alpha(D_{\alpha'}\times B), \tilde{r}_\alpha(D_\alpha\times B)) \\ \le\ & \eta(|\alpha'-\alpha|) + \rho(D_{\alpha'}, D_\alpha) \to 0 \quad \text{as } \alpha' \uparrow \alpha\end{aligned}$$

uniformly with respect to B. (ii) It is similarly proved. *q.e.d.*

Under these assumptions, we can prove the following properties of v_α^* $(\alpha \in [0,1])$.

Theorem 4.2. Suppose that Assumption 2 holds. Let $D_{\alpha'}(\in \mathcal{C}(S)) \downarrow D_\alpha$ as $\alpha' \uparrow \alpha$ for all $\alpha \in (0,1]$. Then

(i) $v^*_{\alpha'}(D_{\alpha'}) \supset v^*_\alpha(D_\alpha)$ for $\alpha' < \alpha$,

(ii) $\lim_{\alpha' \uparrow \alpha} v^*_{\alpha'}(D_{\alpha'}) = v^*_\alpha(D_\alpha)$.

Proof. (i) Let $D_{\alpha'}(\in \mathcal{C}(S)) \downarrow D_\alpha$ as $\alpha' \uparrow \alpha$ for all $\alpha \in (0,1]$. Let $\alpha' < \alpha$ and $v, w \in V$ such that $v(D) \subset w(D')$ for all $D, D' \in \mathcal{C}(S) : D \subset D'$. Let $D, D' \in \mathcal{C}(S) : D \subset D'$. Then

$$\tilde{r}_\alpha(D \times B) + \beta v(\tilde{q}_\alpha(D \times B)) \subset \tilde{r}_{\alpha'}(D' \times B) + \beta w(\tilde{q}_{\alpha'}(D' \times B)) \quad \text{for } B \in \mathcal{C}(A).$$

Therefore

$$U_\alpha v(D) \subset U_{\alpha'} w(D') \quad \text{for } D, D' \in \mathcal{C}(S) : D \subset D'.$$

Thus inductively we obtain

$$(U_\alpha)^t v(D) \subset (U_{\alpha'})^t w(D') \quad \text{for } t = 1, 2, \cdots \text{ and } D, D' \in \mathcal{C}(S) : D \subset D'.$$

Since U_α and $U_{\alpha'}$ are contractive, from Theorem 3.1 we get

$$v^*_\alpha(D_\alpha) \subset v^*_{\alpha'}(D') \quad \text{for } D, D' \in \mathcal{C}(S) : D \subset D'.$$

Therefore

$$v^*_\alpha(D_\alpha) \subset v^*_{\alpha'}(D_{\alpha'}) \quad \text{for } \alpha' < \alpha.$$

(ii) Let $D_{\alpha'}(\in \mathcal{C}(S)) \downarrow D_\alpha$ as $\alpha' \uparrow \alpha$ for all $\alpha \in (0,1]$. We define $J \in V$ by

$$J(D) := [0, M/(1-\beta)] \qquad \text{for all } D \in \mathcal{C}(S).$$

Put $v^{(t)}_\alpha(D) := (U_\alpha)^t J(D)$ for $t = 1, 2, \cdots$ and $D \in \mathcal{C}(S)$. Since U is a contraction map by Theorem 4.1, we have

$$v^{(t)}_\alpha(D) \to v^*_\alpha(D) \quad \text{uniformly for } \alpha \text{ and } D \in \mathcal{C}(S).$$

In order to prove this theorem, it is sufficient to show that, for all $t = 1, 2, \cdots$,

$$\delta(v^{(t)}_{\alpha'}(D_{\alpha'}), v^*_\alpha(D_\alpha)) \to 0 \quad \text{as } \alpha' \uparrow \alpha. \tag{4.3}$$

We show this by induction on t. In the case of $t = 1$, from Lemma 4.1

$$\delta(v^{(1)}_{\alpha'}(D_{\alpha'}), v^{(1)}_\alpha(D_\alpha)) \le \sup_{B \in \mathcal{C}(A)} \delta(\tilde{r}_{\alpha'}(D_{\alpha'}, B), \tilde{r}_\alpha(D_\alpha, B)) \to 0$$

as $\alpha' \uparrow \alpha$. Assuming (4.3) for t, from Lemmas 3.2 and 4.1,

$$\begin{aligned} \delta(v^{(t+1)}_{\alpha'}(D_{\alpha'}), v^{(t+1)}_\alpha(D_\alpha)) \le & \sup_{B \in \mathcal{C}(A)} \delta(\tilde{r}_{\alpha'}(D_{\alpha'}, B), \tilde{r}_\alpha(D_\alpha, B)) \\ & + \beta \sup_{B \in \mathcal{C}(A)} \delta(v^{(t)}_{\alpha'}(\tilde{q}_{\alpha'}(D_{\alpha'}, B)), v^{(t)}_\alpha(\tilde{q}_\alpha(D_\alpha, B))) \to 0 \end{aligned}$$

as $\alpha' \uparrow \alpha$. Thus inductively we get (4.3) for all $t = 1, 2, \cdots$. Therefore we complete the proof. *q.e.d.*

For $\tilde{s} \in \mathcal{F}_c(S)$, define

$$v^*(\tilde{s})(u) := \sup_{\alpha \in [0,1]} \{\alpha \wedge 1_{v^*_\alpha(\tilde{s}_\alpha)}(u)\} \quad u \in [0, M].$$

Then we obtain the following result.

Theorem 4.3. Suppose that Assumption 2 holds. Then

$$v^*(\tilde{s}) \in \mathcal{F}_c([0, M/(1-\beta)]) \quad \text{for all } \tilde{s} \in \mathcal{F}_c(S)$$

and

$$v^*(\tilde{s}) \succeq \psi(\tilde{\pi}, \tilde{s}) \quad \text{for all admissible policies } \tilde{\pi} \text{ and } \tilde{s} \in \mathcal{F}_c(S).$$

Proof. Let $\tilde{s} \in \mathcal{F}_c(S)$. From Theorem 4.2 and [1,Lemma 3], it is trivial that $v^*(\tilde{s}) \in \mathcal{F}_c([0, M/(1-\beta)])$. Next

$$v^*_\alpha(\tilde{s}_\alpha) = U v^*_\alpha(\tilde{s}_\alpha) \succeq_{\mathrm{ci}} U^\pi_\alpha v^*_\alpha(\tilde{s}_\alpha) \quad \text{for } \pi \in \Pi_A.$$

Let $\tilde{\pi} = (\pi_1, \pi_2, \cdots)$ be an admissible Markov policy. Then

$$\psi(\tilde{\pi}, \tilde{s})_\alpha = \psi_\alpha(\tilde{\pi}, \tilde{s}_\alpha) = \lim_{t\to\infty} U^{\pi_1}_\alpha U^{\pi_2}_\alpha \cdots U^{\pi_{t-1}}_\alpha v^*(\tilde{s}_\alpha) \preceq_{\mathrm{ci}} v^*_\alpha(\tilde{s}_\alpha) = v^*(\tilde{s})_\alpha$$

for $\alpha \in [0,1]$. Therefore we obtain that

$$\psi(\tilde{\pi}, \tilde{s}) \preceq v^*(\tilde{s}).$$

q.e.d.

Corollary 4.1. Suppose that there exists $\pi^* \in \Pi_A$ such that $U^{\pi^*}_\alpha v^*_\alpha = v^*_\alpha$ for all $\alpha \in [0,1]$. Then $\pi^{*\infty}$ is absolutely optimal, i.e.,

$$\psi(\pi^{*\infty}, \tilde{s}) \succeq \psi(\tilde{\pi}, \tilde{s}) \quad \text{for all admissible policies } \tilde{\pi} \text{ and } \tilde{s} \in \mathcal{F}_c(S).$$

References.

[1] E.Ammar and J.Metz, On fuzzy convexity and parametric fuzzy optimization, *Fuzzy Sets and Systems*, **49** (1992) 135-141.

[2] R.E.Bellman and L.A.Zadeh, Decision-making in a fuzzy environment, *Manag. Sci., Ser.B*, **17**(1970), 141-164.

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

[4] D.Blackwell, Discrete Dynamic Programming, *Ann. Math. Stat.*, **33** (1962) 719-726.

[5] D.Blackwell, Discounted Dynamic Programming, *Ann. Math. Stat.*, **36** (1965) 226-235.

[6] Chen-wei Xu, On convexity of fuzzy sets and fuzzy relations, *Infor. Sci.*, **59** (1992) 92-102.

[7] A.O.Esogbue and R.E.Bellman, Fuzzy dynamic programming and its extensions, *TIMS/Studies in Manag. Sci.*, **20** *,147-167, North-Holland, Amsterdam, 1984* .

[8] U.D.Holzbaur, A concept of fuzzy transition probabilities, *Ann. Oper. Res.*, **32** (1991) 35-50.

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

[10] L.Madan, 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.Novák, *Fuzzy sets and their Applications*, Adom Hilger, Bristol-Boston, 1989.

[13] 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.

[14] Y.Yoshida, M.Yasuda, J.Nakagami and M.Kurano, A potential of fuzzy relations with a linear structure: The unbounded case. *Submitted.*

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