EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

ELSEVIER European Journal of Operational Research 92 (1996) 649–662

Theory and Methodology

Markov-type fuzzy decision processes with a discounted reward on a closed interval

Masami Kurano [a], Masami Yasuda [b,*], Jun-ichi Nakagami [b], Yuji Yoshida [c]

[a] *Faculty of Education, Chiba University, Chiba 263, Japan*
[b] *Faculty of Science, Chiba University, Chiba 263, Japan*
[c] *Faculty of Economics, Kitakyushu University, Kitakyushu 802, Japan*

Received July 1994; revised May 1995

Abstract

We formulate a new multi-stage decision process with Markov-type fuzzy transition, which is termed Markov-type fuzzy decision process. In the general framework of the decision process, both of state and action are assumed to be fuzzy itself. The transition of states is defined using the fuzzy relation with Markov property and the discounted total reward is described as a fuzzy number on a closed bounded interval. To discuss the optimization problem, a partial order of convex fuzzy numbers is introduced. In this paper the discounted total reward associated with an admissible stationary policy is characterized by a unique fixed point of the contractive mapping. Moreover, the optimality equation for the fuzzy decision model is derived under some continuity conditions. Also, an illustrated example is given to explain the theoretical results and the computation in the paper.

Keywords: Fuzzy decision process; Markov-type transition; Partial order; Policy improvement; Convex fuzzy relation

1. Introduction

Fuzzy decision making, introduced by Bellman and Zadeh [3], is a multi-stage process in which the goals and/or the constraints are fuzzy. The method of dynamic programming is shown to be a powerful computational technique for these problems [3,8]. For a wide application, it is desirable to develop a theory for the multi-stage decision process under fuzzification of the state and its transition by fuzzy relations. Baldwin and Pilsworth [2] have proposed a multi-stage decision model described by fuzzy mappings which is an extension of the Bellman–Zadeh model. They have taken an optimal decision which maximizes the measure of the truthness of "control and goal constraints satisfied".

In this paper, by trying to do direct fuzzification of the deterministic decision system, we will define a multi-stage decision process with Markov-type fuzzy transition [10,14,15], which is different from the Baldwin–Pilsworth model. The optimization of the discounted total reward for the processes under some partial

* Corresponding author.

SSDI 0377-2217(95)00140-9

order, called "fuzzy max order", on the class of convex fuzzy numbers is considered. Based on the α-cut of the fuzzy set the analysis is done by operators on some class of functions, which is used in Markov decision processes (for example, see [5,6]) and, applying Banach's fixed point theorem, the discounted total fuzzy reward from any fuzzy policy satisfying some reasonable conditions is obtained as a unique solution of the related fuzzy relational equations. Also, an optimality fuzzy relational equation is given to characterize an optimal fuzzy policy.

In Section 2, we list the notations and construct the model to be analyzed in the succeeding sections. In Section 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. A numerical example is given in Section 5.

2. The fuzzy decision model

In this section, we shall give notations and mathematical facts in order to describe a fuzzy decision processes considered in the sequel. Let E, E_1, E_2 be convex 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 [16] and Novák [13] for the theory of fuzzy sets.

The set of all fuzzy sets $\tilde{s}$ on E is denoted by $\mathscr{F}(E)$, which are assumed, throughout the paper, to be upper semi-continuous and have a compact support with the normality condition: $\sup_{x\in E}\tilde{s}(x)=1$.

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

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

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

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

where $a\wedge b=\min\{a, b\}$ [1]. Note that $\tilde{s}$ is convex iff the α-cut $\tilde{s}_\alpha$ is a convex set for all $\alpha\in[0, 1]$ [7]. Some papers on convex analysis call this notion quasi-concave.

A fuzzy relation $\tilde{p}\in\mathscr{F}(E_1\times E_2)$ 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 subscript c as

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

The set of all non-empty convex closed subsets of E is denoted by $\mathscr{C}(E)$. Then clearly $\tilde{s}\in\mathscr{F}_c(E)$ means that $\tilde{s}_\alpha\in\mathscr{C}(E)$ for all $\alpha\in[0, 1]$.

Let us restrict the term of convex fuzzy number to be that of the closed convex support contained in the interval $[0, M]\subset\mathbb{R}_+:=[0, \infty]$ with a positive number M, that is,

$$\mathscr{F}_c([0, M]):=\{\tilde{s}\in\mathscr{F}_c(\mathbb{R}_+)\mid\tilde{s}_0\subset[0, M]\},$$

and let $\mathscr{C}([0, M])$ be the set of all closed intervals of $[0, M]$. For non-empty closed intervals, the Hausdorff metric δ can be considered and it becomes a complete separable metric space, i.e.,

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

where $a \vee b = \max\{a, b\}$. The addition and the scalar multiplication of fuzzy sets on $\mathscr{F}(\mathbb{R}_+)$ are defined as follows [11]: For $\tilde{n}, \tilde{m} \in \mathscr{F}(\mathbb{R}_+)$ and $\lambda \in \mathbb{R}_+$, define

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

and

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

where $I_A(\cdot)$ is the classical indicator function of a set $A \subset \mathbb{R}_+$. 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 for these operations. Here the operation on sets is defined ordinary as $A + B := \{x + y \mid x \in A,\ y \in B\}$ and $\lambda A := \{\lambda x \mid x \in A\}$ for $A, B \subset \mathbb{R}_+$.

The following result appeared in [7]:

Lemma 2.1 ([7]).
(i) *For any* $\tilde{n}, \tilde{m} \in \mathscr{F}_c(\mathbb{R}_+)$ *and* $\lambda \in \mathbb{R}_+$, $\tilde{n} + \tilde{m} \in \mathscr{F}_c(\mathbb{R}_+)$ *and* $\lambda \tilde{n} \in \mathscr{F}_c(\mathbb{R}_+)$.
(ii) *For any* $\tilde{s} \in \mathscr{F}_c(E_1)$ *and* $\tilde{p} \in \mathscr{F}_c(E_1 \times E_2)$, *then* $\sup_{x \in E_1} \tilde{s}(x) \wedge \tilde{p}(x, \cdot) \in \mathscr{F}_c(E_2)$.

Now, we consider Markov-type fuzzy decision processes constructed by six objects $(S, A, [0, M], \tilde{q}, \tilde{r}, \beta)$, which satisfy the following conditions:

(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 elements of $\mathscr{F}_c(S)$ and $\mathscr{F}_c(A)$, 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 $\tilde{q} \in \mathscr{F}_c(S \times A \times S)$ and $\tilde{r} \in \mathscr{F}_c(S \times A \times [0, M])$, where M is a given positive number. Explicitly, if the system is in a fuzzy state $\tilde{s} \in \mathscr{F}_c(S)$ and the fuzzy action $\tilde{a} \in \mathscr{F}_c(A)$ 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 obtained, where Q and R are defined by the following:

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

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

Note that, by Lemma 2.1, it holds that $Q(\tilde{s}, \tilde{a})(\cdot) \in \mathscr{F}_c(S)$ and $R(\tilde{s}, \tilde{a})(\cdot) \in \mathscr{F}_c([0, M])$ for all $\tilde{s} \in \mathscr{F}_c(S)$, $\tilde{a} \in \mathscr{F}_c(A)$.

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

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

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

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

The transition of fuzzy states by (2.3) has the Markov property, that is, the state of the $(t+1)$th step is determined by that of tth 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 from a fuzzy policy $\check{\pi}$, let us consider the convergence of a sequence of fuzzy numbers belonging to $\mathscr{F}_c(\mathbb{R}_+)$.

Definition 2.1 ([10, 12]). For $\tilde{n}_t, \tilde{n} \in \mathscr{F}_c(\mathbb{R}_+)$, $\lim_{t\to\infty} \tilde{n}_t = \tilde{n}$ 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 the convergence theorem proved in [14].

Lemma 2.2. *For $\tilde{s} \in \mathscr{F}_c(S)$ and $\check{\pi} = (\pi_1, \pi_2, \ldots)$,*

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

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

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

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

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

The problem is to maximize the fuzzy reward $\psi(\check{\pi}, \tilde{s})$ over a certain class of fuzzy policies $\{\check{\pi}\}$ with respect to a given partial order on $\mathscr{F}_c([0, M/(1-\beta)])$. 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 [16], which gives a natural extension of non-fuzzy systems. To explain the notion, we treat the following usual continuous deterministic systems: For a given $a_t \in A$, $t = 1, 2, \ldots$, the transition of states is described by

$$x_{t+1} = f(x_t, a_t) \quad (t = 1, 2, \ldots), \tag{2.6}$$

where $x_1 \in S$ is an initial state and $f: S \times A \mapsto S$ is a continuous deterministic transition function. A fuzzy relation $\tilde{f} \in \mathscr{F}(S \times A \times S)$ 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 (2.6) can be rewritten in the fuzzy environments:

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

where $\tilde{x}_t = 1_{\{x_t\}}$, $\tilde{a}_t = 1_{\{a_t\}}$ and

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

This shows that the transition (2.1) of fuzzy states is a fuzzy extension of the deterministic system (2.6) by extending the state space S and action space A to fuzzy sets $\mathscr{F}(S)$ and $\mathscr{F}(A)$, respectively.

3. Partial order for the optimization

We introduce a partial order on $\mathscr{F}_c([0, M/(1-\beta)])$ and give some results on the optimization of the fuzzy decision processes defined in the previous section. For $\tilde{n}, \tilde{m} \in \mathscr{F}_c([0, M/(1-\beta)])$, a partial order for fuzzy numbers is defined as

$$\tilde{n} \succcurlyeq \tilde{m}$$

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 and max mean the left or right end point of the α-cut interval respectively [9]. It is immediate that $(\mathscr{F}_c([0, M/(1-\beta)]), \succcurlyeq)$ becomes a complete lattice [4]. Note also that

$$\sup_t \tilde{u}_t \in \mathscr{F}_c([0, M/(1-\beta)]) \quad \text{for } \{\tilde{u}_t\} \subset \mathscr{F}_c([0, M/(1-\beta)])$$

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

Definition 3.1. The fuzzy strategy $\pi : \mathscr{F}_c(S) \mapsto \mathscr{F}_c(A)$ is called admissible if the α-cut set $\pi(\tilde{s})_\alpha$ of π depends only on the scalar α and the set $\tilde{s}_\alpha$, that is, it can 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 $\check{\pi} = (\pi_1, \pi_2, \ldots)$ is called admissible if $\pi_t \in \Pi_A$ $(t = 1, 2, \ldots)$.

Our problem is to maximize $\psi(\check{\pi}, \tilde{s})$ over all admissible policies $\check{\pi}$ with respect to the order $\succcurlyeq$ on $\mathscr{F}_c([0, M/(1-\beta)])$.

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

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

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

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

By using $\tilde{q}_\alpha$ and $\tilde{r}_\alpha$, define maps $Q_\alpha^\pi : \mathscr{C}(S) \mapsto \mathscr{C}(S)$ and $R_\alpha^\pi : \mathscr{C}(S) \mapsto \mathscr{C}([0, M])$ $(\pi \in \Pi_A, \alpha \in [0, 1])$ by

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

for $D \in \mathscr{C}(S)$. For any admissible fuzzy policy $\check{\pi} = (\pi_1, \pi_2, \ldots)$, $Q_{t,\alpha}^{\check{\pi}}$ $(t \geq 0)$ is defined inductively by using the composition of maps as follows:

$$Q_{0,\alpha}^{\check{\pi}} := I \text{ (identity)}, \qquad Q_{1,\alpha}^{\check{\pi}}(D) := Q_\alpha^{\pi_1}(D), \qquad Q_{t+1,\alpha}^{\check{\pi}}(D) := Q_\alpha^{\pi_{t+1}} Q_{t,\alpha}^{\check{\pi}}(D),$$

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

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

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

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

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

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

Proof. (i) From Lemma 1 in [10], we see that $Q(\tilde{u}, \pi(\tilde{u}))_\alpha = Q^\pi_\alpha(\tilde{u}_\alpha)$ and $R(\tilde{u}, \pi(\tilde{u}))_\alpha = R^\pi_\alpha(\tilde{u}_\alpha)$ for all $\tilde{u} \in \mathscr{F}_c(S)$ and $\pi \in \Pi_A$. Based on this fact, (i) is proved by induction. For $t = 1$, we have

$$\tilde{s}_{2,\alpha} = Q(\tilde{s}, \pi_1(\tilde{s}))_\alpha = Q^{\pi_1}_\alpha(\tilde{s}_\alpha) = Q^{\check{\pi}}_{1,\alpha}(\tilde{s}_\alpha).$$

Assume (i) for some t ($t \geq 1$). Then

$$\tilde{s}_{t+1,\alpha} = Q(\tilde{s}_t, \pi_t(\tilde{s}_t))_\alpha = Q^{\pi_t}_\alpha(\tilde{s}_{t,\alpha}) = Q^{\pi_t}_\alpha Q^{\check{\pi}}_{t-1,\alpha}(\tilde{s}_\alpha) = Q^{\check{\pi}}_{t,\alpha}(\tilde{s}_\alpha),$$

which implies that (i) holds for $t + 1$. Also, (ii) holds obviously. (iii) follows immediately from the property of α-cuts of the fuzzy numbers and Lemma 2.2. □

Let $V := \{v : \mathscr{C}(S) \mapsto \mathscr{C}([0, M/(1-\beta)])\}$. Define a metric d_V on V by

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

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

$$v \succcurlyeq_V w$$

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

$$U^\pi_\alpha v(D) := R^\pi_\alpha(D) + \beta v(Q^\pi_\alpha(D)) \tag{3.2}$$

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

We will prove the contractive property of the operator U^π_α. To do it, we need the following lemma whose proof is easy.

Lemma 3.2.
(i) *Let $\Gamma := \{\gamma\}$ be an index set and $[a_\gamma, b_\gamma], [c_\gamma, d_\gamma] \in \mathscr{C}([0, M/(1-\beta)])$ for $\gamma \in \Gamma$. Then*

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

(ii) *If $[a_1, b_1], [c_1, d_1], [a_2, b_2], [c_2, d_2] \in \mathscr{C}([0, M/(1-\beta)])$, then*

$$\delta([a_1, b_1] + [c_1, d_1], [a_2, b_2] + [c_2, d_2]) \leq \delta([a_1, b_1], [a_2, b_2]) + \delta([c_1, d_1], [c_2, d_2]).$$

(iii) *If $[a, b], [c, d] \in \mathscr{C}([0, M/(1-\beta)])$, then*

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

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. Let $v, w \in V$ be two maps such that

$$U_\alpha^\pi v(D) = R_\alpha^\pi(D) + \beta v(Q_\alpha^\pi(D)) \quad \text{and} \quad U_\alpha^\pi w(D) = R_\alpha^\pi(D) + \beta w(Q_\alpha^\pi(D)).$$

By using Lemma 3.2, we have

$$\begin{aligned}\delta(U_\alpha^\pi v(D), U_\alpha^\pi w(D)) &\le \delta(R_\alpha^\pi(D)), R_\alpha^\pi(D)) + \delta(\beta v(Q_\alpha^\pi(D)), \beta w(Q_\alpha^\pi(D)))\\ &= \beta\delta(v(Q_\alpha^\pi(D)), w(Q_\alpha^\pi(D))) \le \beta d_V(v, w)\end{aligned}$$

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

$$d_V(U_\alpha^\pi v, U_\alpha^\pi w) \le \beta d_V(v, w).$$

That is, U_α^π is contractive and, by Banach's fixed point theorem, has a unique map $v_\alpha^\pi \in V$ such that $v_\alpha^\pi = U_\alpha^\pi v_\alpha^\pi$. Further if $v \succcurlyeq_V w$, then we have

$$U_\alpha^\pi v(D) = R_\alpha^\pi(D) + \beta v(Q_\alpha^\pi(D)) \succcurlyeq_{\mathrm{ci}} R_\alpha^\pi(D) + \beta w(Q_\alpha^\pi(D)) = U_\alpha^\pi w(D)$$

for all $D \in \mathscr{C}(S)$. So $U_\alpha^\pi v \succcurlyeq_V U_\alpha^\pi w$. Therefore U_α^π is monotone. □

The discounted total reward associated with the admissible stationary policy is characterized as follows.

Theorem 3.2. *For $\tilde{s} \in \mathscr{F}_c(S)$ and any admissible stationary policy $\pi^\infty = (\pi, \pi, \pi, \dots)$,*

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

holds for $\alpha \in [0, 1]$, where $v_\alpha^\pi \in V$ is a unique fixed point of the contractive map U_α^π.

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

$$\psi_\alpha(\pi^\infty, D) := \sum_{t=1}^{\infty} \beta^{t-1} R_\alpha^\pi(Q_{t-1,\alpha}^{\pi^\infty}(D))$$

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

$$\psi(\pi^\infty, \tilde{s})_\alpha = \sum_{t=1}^{\infty} \beta^{t-1} R(\tilde{s}_t, \pi(\tilde{s}_t))_\alpha = \sum_{t=1}^{\infty} \beta^{t-1} R_\alpha^\pi(\tilde{s}_{t,\alpha}) = \psi_\alpha(\pi^\infty, \tilde{s}_\alpha).$$

On the otherhand, we have

$$\begin{aligned}\psi_\alpha(\pi^\infty, D) &= R_\alpha^\pi(D) + \beta\left\{\sum_{t=2}^{\infty} \beta^{t-2} R_\alpha^\pi(Q_{t-2,\alpha}^{\pi^\infty}(Q_\alpha^\pi(D)))\right\}\\ &= R_\alpha^\pi(D) + \beta\psi_\alpha(\pi^\infty, Q_\alpha^\pi(D)) = U_\alpha^\pi \psi_\alpha(\pi^\infty, \cdot)(D)\end{aligned}$$

for any $D \in \mathscr{C}(S)$. Therefore $\psi_\alpha(\pi^\infty, \cdot)$ is a fixed point of U_α^π. From Theorem 3.1, $v_\alpha^\pi(\cdot) = \psi_\alpha(\pi^\infty, \cdot)$, which implies that $(\psi(\pi^\infty, \tilde{s}))_\alpha = \psi_\alpha(\pi^\infty, \tilde{s}_\alpha) = v_\alpha^\pi(\tilde{s}_\alpha)$. □

For any admissible policy $\check{\pi} = (\pi_1, \pi_2, \dots)$, let

$$\psi_\alpha(\check{\pi}, D) = \sum_{t=1}^{\infty} \beta^{t-1} R_\alpha^{\pi_t}(Q_{t-1,\alpha}^{\check{\pi}}(D)) \quad (D \in \mathscr{C}(\mathscr{S})).$$

Then, as in the proof of Theorem 3.2, we have $(\psi(\check{\pi}, \tilde{s}))_\alpha = \psi_\alpha(\check{\pi}, \tilde{s}_\alpha)$ $(\alpha \in [0, 1])$. The following lemma shows that $\psi_\alpha(\check{\pi}, \tilde{s}_\alpha)$ can be represented by using the operator $U_\alpha^{\pi_t}$, $t \ge 1$.

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

$$\psi_\alpha(\check{\pi}, \tilde{s}_\alpha) = U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} \psi_\alpha(\check{\pi}_t, \cdot)(\tilde{s}_\alpha)$$

where $\check{\pi}_t = (\pi_t, \pi_{t+1}, \pi_{t+2}, \dots)$.
(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) By observing the proof of Theorem 3.2, it holds

$$\psi_\alpha(\check{\pi}, \tilde{s}_\alpha) = U_\alpha^{\pi_1} \psi_\alpha(\check{\pi}_1, \cdot)(\tilde{s}_\alpha).$$

So, (i) is proved inductively for $t \geq 2$.
(ii) Similarly as Theorem 3.1 we have

$$\delta\left(U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} \psi_\alpha(\check{\pi}_t, \cdot)(\tilde{s}_\alpha), U_\alpha^{\pi_1} U_\alpha^{\pi_2} \cdots U_\alpha^{\pi_{t-1}} v(\tilde{s}_\alpha)\right) \leq \beta^{t-1} \delta\left(\psi_\alpha(\check{\pi}_t, \tilde{s}_\alpha), v(\tilde{s}_\alpha)\right)$$

which yields (ii) by $t \to \infty$. □

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

$$\psi_\alpha(\check{\pi}, D) \succcurlyeq_{ci} U_\alpha^{\pi} \psi_\alpha(\check{\pi}, \cdot)(D) \tag{3.4}$$

for all $D \in \mathscr{C}(S)$, $\pi \in \Pi_A$ and $\alpha \in [0, 1]$. Then we have

$$\psi(\check{\pi}, \tilde{s}) \succcurlyeq \psi(\check{\sigma}, \tilde{s}) \tag{3.5}$$

for all $\tilde{s} \in \mathscr{F}_c(S)$ and any admissible policy $\check{\sigma}$.

Proof. By Lemma 3.3 and the monotonicity of U_α^{π}, it can be shown that

$$\psi_\alpha(\check{\pi}, \tilde{s}_\alpha) \succcurlyeq_{ci} \psi_\alpha(\check{\sigma}, \tilde{s}_\alpha)$$

for all $\alpha \in [0, 1]$ and any admissible policy $\check{\sigma}$. This implies the result easily. □

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

$$U_\alpha^{\pi} \psi_\alpha(\check{\pi}, \cdot)(D) \succcurlyeq_{ci} \psi_\alpha(\check{\pi}, D) \tag{3.6}$$

for all $D \in \mathscr{C}(S)$ and $\alpha \in [0, 1]$. Then we have

$$\psi(\pi^\infty, \tilde{s}) \succcurlyeq \psi(\check{\pi}, \tilde{s}) \tag{3.7}$$

for all $\tilde{s} \in \mathscr{F}_c(S)$.

Proof. Similarly as the proof of Theorem 3.3, it can be shown easily. □

Remark. Results like Theorem 3.3 and 3.4 have already appeared in the cl sic discounted Markov decision model and used for the policy improvement [5,6]. By the same idea, the at 'e theorems would be useful in 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

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

for $v \in V$ and $D \in \mathscr{C}(S)$, where the supremum is taken with respect to the order $\succcurlyeq_{ci}$.

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 3.2, for $v, w \in V$ we obtain

$$\delta(U_\alpha v(D), U_\alpha w(D)) \le \sup_{B \in \mathscr{C}(A)} \delta(\beta v(\tilde{q}_\alpha(D \times B), \beta w(\tilde{q}_\alpha(D \times B))$$

$$= \beta \sup_{B \in \mathscr{C}(A)} \delta(v(\tilde{q}_\alpha(D \times B), w(\tilde{q}_\alpha(D \times B)) \le \beta d_V(v, w)$$

for $D \in \mathscr{C}(S)$. Therefore $d_V(U_\alpha v, U_\alpha w) \le \beta d_V(v, w)$. By Banach's fixed point theorem, there exists a unique $v_\alpha^* \in V$ such that $v_\alpha^* = U_\alpha v_\alpha^*$. Also the monotonicity of U_α follows obviously. □

It will be shown in this section that a unique fixed point of U_α gives the α-cut of the maximum fuzzy reward under some continuity conditions, so that (4.2) is interpreted as the optimality equation for our fuzzy decision model.

For any $(x, a) \in S \times A$ and $\alpha \in [0, 1]$, let

$$\tilde{r}_\alpha(x, a) := \tilde{r}_\alpha(\{x\} \times \{a\}) \quad \text{and} \quad \tilde{q}_\alpha(x, a) := \tilde{q}_\alpha(\{x\} \times \{a\}).$$

Then, by the definition we have, for each $D \in \mathscr{C}(S)$ and $B \in \mathscr{C}(A)$,

$$\tilde{r}_\alpha(D \times B) := \bigcup_{(x,a) \in D \times B} \tilde{r}_\alpha(x, a) \quad \text{and} \quad \tilde{q}_\alpha(D \times B) := \bigcup_{(x,a) \in D \times B} \tilde{q}_\alpha(x, a).$$

Condition A. (A uniform continuity on $\tilde{r}$ and $\tilde{q}$). There exists non-decreasing and continuous η: $[0, \infty) \mapsto [0, \infty)$ such that

(i) $\eta(t) \to 0$ as $t \downarrow 0$;
(ii) $\delta(\tilde{r}_{\alpha'}(x, a), \tilde{r}_\alpha(x, a)) \le \eta(|\alpha' - \alpha|)$ for $0 \le \alpha' < \alpha$;
(iii) $\rho_S(\tilde{q}_{\alpha'}(x, a), \tilde{q}_\alpha(x, a)) \le \eta(|\alpha' - \alpha|)$ for $0 \le \alpha' < \alpha$;
(iv) $\delta(\tilde{r}_\alpha(x', a), \tilde{r}_\alpha(x, a)) \le \eta(d_S(x', x))$ for $x, x' \in S$, $a \in A$, $\alpha \in (0, 1]$;
(v) $\rho_S(\tilde{q}_\alpha(x', a), \tilde{q}_\alpha(x, a)) \le \eta(d_S(x', x))$ for $x, x' \in S$, $a \in A$, $\alpha \in (0, 1]$,

where ρ_S is the Hausedorff metric on $\mathscr{C}(S)$ induced by a metric d_S on S.

Lemma 4.1. *Suppose that Condition A holds. Let $D_{\alpha'} (\in \mathscr{C}(S)) \downarrow D_\alpha$ as $\alpha' \uparrow \alpha$ for all $\alpha \in (0, 1]$. Then*
(i) $\sup_{B \in \mathscr{C}(A)} \delta(\tilde{r}_{\alpha'}(D_{\alpha'} \times B), \tilde{r}_\alpha(D_\alpha \times B)) \to 0$ *as* $\alpha' \uparrow \alpha$;
(ii) $\sup_{B \in \mathscr{C}(A)} \rho_S(\tilde{q}_{\alpha'}(D_{\alpha'} \times B), \tilde{q}_\alpha(D_\alpha \times B)) \to 0$ *as* $\alpha' \uparrow \alpha$.

Proof. Let $0 \le \alpha' \le \alpha \le 1$. Then, since

$$\delta(\tilde{r}_{\alpha'}(D_{\alpha'} \times B), \tilde{r}_{\alpha}(D_{\alpha'} \times B)) = \delta\left(\bigcup_{(x,a) \in D_{\alpha'} \times B} \tilde{r}_{\alpha'}(x, a), \bigcup_{(x,a) \in D_{\alpha'} \times B} \tilde{r}_{\alpha}(x, a) \right)$$
$$\le \sup_{(x,a) \in D_{\alpha'} \times B} \delta(\tilde{r}_{\alpha'}(x, a), \tilde{r}_{\alpha}(x, a)), \tag{4.3}$$

we have

$$\delta(\tilde{r}_{\alpha'}(D_{\alpha'} \times B), \tilde{r}_{\alpha}(D_{\alpha'} \times B)) \le \eta(|\alpha' - \alpha|).$$

On the other hand, from Condition A,

$$\delta(\tilde{r}_{\alpha}(D_{\alpha'} \times B), \tilde{r}_{\alpha}(D_{\alpha} \times B)) \le \max_{(x',a) \in D_{\alpha'} \times B} \min_{x \in D_{\alpha}} \delta(\tilde{r}_{\alpha}(x', a), \tilde{r}_{\alpha}(x, a))$$
$$\le \max_{x' \in D_{\alpha'}} \min_{x \in D_{\alpha}} \eta(d_S(x', s)) = \max_{x' \in D_{\alpha'}} \eta(d_S(x', D_{\alpha}))$$
$$\le \eta(\rho_S(D_{\alpha'}, D_{\alpha})). \tag{4.4}$$

Thus we get

$$\delta(\tilde{r}_{\alpha'}(D_{\alpha'} \times B), \tilde{r}_{\alpha}(D_{\alpha} \times B))$$
$$= \delta(\tilde{r}_{\alpha'}(D_{\alpha'} \times B), \tilde{r}_{\alpha}(D_{\alpha'} \times B)) + \delta(\tilde{r}_{\alpha}(D_{\alpha'} \times B), \tilde{r}_{\alpha}(D_{\alpha} \times B))$$
$$\le \eta(|\alpha' - \alpha|) + \eta(\rho_S(D_{\alpha'}, D_{\alpha})) \quad \text{as } \alpha' \uparrow \alpha$$

uniformly with respect to B. Also, (4.3) and (4.4) holds for $\tilde{q}_{\alpha}$, ρ_S, so (ii) can be checked similarly. □

Some properties of v_{α}^* ($\alpha \in [0, 1]$) are investigated in the following theorem.

Theorem 4.2. *Suppose that Condition A holds. Let $D_{\alpha'}(\in \mathscr{C}(S)) \downarrow D_{\alpha}$ as $\alpha' \uparrow \alpha$ for $\alpha \in (0, 1]$. 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. Let $\alpha' < \alpha$ and $v, w \in V$ such that $v(D) \subset w(D')$ for all $D, D' \in \mathscr{C}(S)$ with $D \subset D'$. Then, for $D, D' \in \mathscr{C}(S)$ with $D \subset D'$, we have

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

Therefore, by (4.1),

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

Thus inductively we obtain

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

Since U_{α} and $U_{\alpha'}$ are contractive, $(U_{\alpha})^t v \to v_{\alpha}^*$ and $(U_{\alpha'})^t w \to v_{\alpha'}^*$ as $t \to \infty$, so that

$$v_{\alpha}^*(D) \subset v_{\alpha'}^*(D'),$$

which implies (i).

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

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

Put $v_\alpha^{(t)}(D) := (U_\alpha)^t J(D)$ for $t = 1, 2, \dots$ and $D \in \mathscr{C}(S)$. Since U_α is a contraction map with modulus β, we have

$$v_\alpha^{(t)}(D) \to v_\alpha^*(D) \quad \text{uniformity for } \alpha \text{ and } D \in \mathscr{C}(S) \quad \text{as } t \to \infty.$$

Therefore in order to prove (ii), it is sufficient to show: For all $t = 1, 2, \dots,$

$$\delta\left(v_{\alpha'}^{(t)}(D_{\alpha'}), v_\alpha^{(t)}(D_\alpha)\right) \to 0 \quad \text{as } \alpha' \uparrow \alpha. \tag{4.5}$$

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

$$\delta\left(v_{\alpha'}^{(1)}(D_{\alpha'}), v_\alpha^{(1)}(D_\alpha)'\right) \le \sup_{B \in \mathscr{C}(A)} \delta(\tilde{r}_{\alpha'}(D_{\alpha'} \times B), \tilde{r}_\alpha(D_\alpha \times B)) \to 0$$

as $\alpha' \uparrow \alpha$. Assuming (4.5) for t, from Lemmas 3.2 and 4.1

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

as $\alpha' \uparrow \alpha$. Thus (4.5) holds for all $t = 1, 2, \cdots$ and we complete the proof. □

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

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

Then we obtain the following result.

Theorem 4.3. *Suppose that Condition A holds. Then*

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

and

$$v^*(\tilde{s}) \succcurlyeq \psi(\check{\pi}, \tilde{s}) \quad \textit{for all admissible policies } \check{\pi} \textit{ and } \tilde{s} \in \mathscr{F}_c(S).$$

Proof. Let $\tilde{s} \in \mathscr{F}_c(S)$. From Theorem 4.2 and [10,Lemma 3], it is trivial that $v^*(\tilde{s}) \in \mathscr{F}_c([0, M/(1-\beta)])$. It follows from (4.1) that

$$v_\alpha^*(\tilde{s}_\alpha) = U v_\alpha^*(\tilde{s}_\alpha) \succcurlyeq_{ci} U_\alpha^\pi v_\alpha^*(\tilde{s}_\alpha) \quad \text{for } \pi \in \Pi_A.$$

Let $\check{\pi} = (\pi_1, \pi_2, \cdots)$ be an admissible policy. Then, by Lemma 3.3(ii),

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

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

$$\psi(\check{\pi}, \tilde{s}) \preccurlyeq v^*(\tilde{s}). \quad \square$$

Corollary 4.1. *Suppose that Condition A holds. If there exists $\pi^* \in \Pi_A$ such that $U_\alpha^{\pi^*} v_\alpha^* = v_\alpha^*$ for all $\alpha \in [0, 1]$, then $\pi^{*\infty}$ is absolutely optimal, i.e.,*

$$\psi(\pi^{*\infty}, \tilde{s}) \succcurlyeq \psi(\check{\pi}, \tilde{s}) \quad \textit{for all admissible policies } \check{\pi} \textit{ and } \tilde{s} \in \mathscr{F}_c(S).$$

Proof. From the assumption of Corollary 4.1, we get inductively

$$\left(U_\alpha^{\pi^*}\right)^t v_\alpha^* = v_\alpha^* \quad \text{for all } t \ge 1.$$

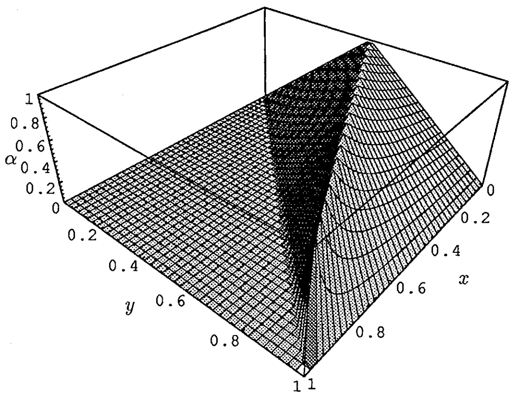


Fig. 1. The fuzzy relation $\tilde{q}(x, a, y)$ when $a = 1$.

As $t \to \infty$ in the above, we have

$$\psi_\alpha(\pi^{*\infty}, \tilde{s}) = v_\alpha^*$$

by Lemma 3.3(ii). Hence $\pi^{*\infty}$ is absolutely optimal from Theorem 4.3. □

5. A numerical example

In this section we give a numerical example to illustrate the theoretical results and computation in the preceding sections. Let $S := [0, 1]$, $A := [0, 1]$, $M := 1$ and $0 < \beta < 1$. The fuzzy relation and the reward are given by

$$\begin{aligned} \tilde{q}(x, a, y) &= \Lambda_{[x \wedge a, x \vee a]}(y), \quad x, y \in S,\ a \in A; \\ \tilde{r}(x, a, r) &= \Lambda_{[x \wedge a,\, x \vee a]}(r), \quad x \in S,\ a \in A,\ r \in [0, 1], \end{aligned} \tag{5.1}$$

respectively, where

$$\Lambda_{[c,d]}(z) = \begin{cases} \left(\left(1 - \dfrac{2}{d-c}\left|\dfrac{c+d}{2} - z\right|\right)\right) \vee 0, & 0 \le c < d \le 1,\ 0 \le z \le 1, \\ \mathbf{1}_{\{c\}}(z), & c = d. \end{cases} \tag{5.2}$$

Fig. 1 shows the fuzzy relation $\tilde{q}(x, \cdot, y)$ when $a = 1$. Clearly Condition A holds with $\eta(x) = x (x \ge 0)$. We shall now investigate an admissible policy $\check{\pi}$ of taking action 1, that is, $\pi = \pi^\infty = (\pi, \pi, \dots)$ and $\pi(\tilde{s}) = 1_{\{1\}}$ for any fuzzy state $\tilde{s} \in \mathscr{F}_c([0, 1])$.

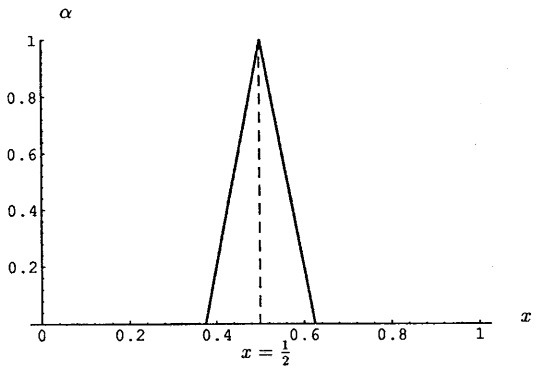
First the discounted total fuzzy reward $\psi(\check{\pi}, \cdot)$, whose α-cut is solved by a unique solution of the following equation:

$$v = U_\alpha^\pi v \quad (v \in V), \tag{5.3}$$

which is given in Theorem 3.1.

Let $[a, b] \subset [0, 1]$ and $0 \le \alpha \le 1$. Observing that

$$\{y \in [0, 1] \mid \Lambda_{[x,1]}(y) \ge \alpha \text{ for some } x \in [a, b]\} = \left[\left(1 - \tfrac{1}{2}\alpha\right)a + \tfrac{1}{2}\alpha,\ \tfrac{1}{2}\alpha b + 1 - \tfrac{1}{2}\alpha\right],$$



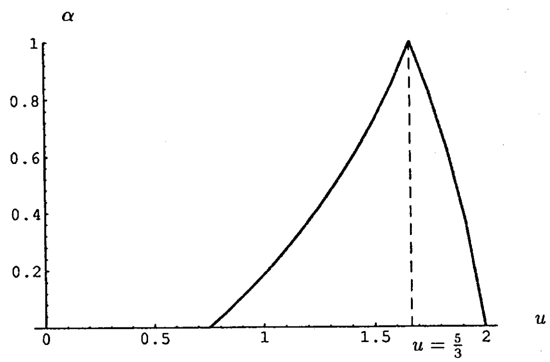


Fig. 2. The initial fuzzy state $\tilde{s}_1(x) = \tilde{s}(x)$.

Fig. 3. The optimal fuzzy rewards $\psi(\check{\pi}, \tilde{s})(u)$.

we easily have

$$Q_\alpha^\pi([a, b]) = R_\alpha^\pi([a, b]) = \left[\left(1 - \tfrac{1}{2}\alpha\right)a + \tfrac{1}{2}\alpha,\ \tfrac{1}{2}\alpha b + 1 - \tfrac{1}{2}\alpha\right].$$

From (3.2),

$$U_\alpha^\pi v([a, b]) = \left[\left(1 - \tfrac{1}{2}\alpha\right)a + \tfrac{1}{2}\alpha,\ \tfrac{1}{2}\alpha b + 1 - \tfrac{1}{2}\alpha\right] + \beta v\left(\left[\left(1 - \tfrac{1}{2}\alpha\right)a + \tfrac{1}{2}\alpha,\ \tfrac{1}{2}\alpha b + 1 - \tfrac{1}{2}\alpha\right]\right). \tag{5.4}$$

In order to calculate the interval equation derived from (5.3) and (5.4), denote the closed interval by $[T_1(a), T_2(b)] := v([a, b])$. Then

$$\begin{aligned} T_1(a) &= \left(1 - \tfrac{1}{2}\alpha\right)a + \tfrac{1}{2}\alpha + \beta T_1\left(\left(1 - \tfrac{1}{2}\alpha\right)a + \tfrac{1}{2}\alpha\right), \\ T_2(b) &= \left(\tfrac{1}{2}\alpha\right)b + 1 - \tfrac{1}{2}\alpha + \beta T_1\left(\tfrac{1}{2}\alpha b + 1 - \tfrac{1}{2}\alpha\right). \end{aligned} \tag{5.5}$$

The above equation (5.5) is solved by the contractive property of the map defined in the following. If $T : [0, 1] \mapsto [0, 1/(1-\beta)]$ is a continuous function satisfying $T(x) = rx + c + \beta T(rx + c)$, where $0 \le r + c \le 1$ and $0 \le c \le 1$, then $T(x)$ is uniquely determined by $T(x) = \{r/(1 - \beta r)\}x + c/\{(1-\beta)(1-\beta r)\}$.

Therefore the unique solution of (5.5) is given

$$T_1(a) = \frac{2-\alpha}{2-(2-\alpha)\beta}a + \frac{\alpha}{(1-\beta)(2-\beta(2-\alpha))}, \qquad T_2(b) = \frac{\alpha}{2-\alpha\beta}b + \frac{2-\alpha}{(1-\beta)(2-\alpha\beta)} \tag{5.6}$$

By Theorem 3.2, we get

$$\psi(\check{\pi}, \tilde{s})_\alpha = \left[T_1(\min \tilde{s}_\alpha), T_2(\min \tilde{s}_\alpha)\right]$$

for the closed interval $[\min \tilde{s}_\alpha, \max \tilde{s}_\alpha]$ with the α-cut of $\tilde{s}$. Also

$$\psi(\check{\pi}, \tilde{s})(r) = \sup_{\alpha \in [0,1]} \left\{\alpha \wedge 1_{\psi(\check{\pi}, \tilde{s})_\alpha}(r)\right\} \tag{5.7}$$

for $0 \le r \le 1/(1-\beta)$.

As a numerical example, let the initial fuzzy state be

$$\tilde{s}_1(x) = \tilde{s}(x) := (1 - |8x - 4|) \vee 0, \quad x \in S$$

and $\beta = 0.5$.

Then, $\psi(\check{\pi}, \tilde{s})$ becomes

$$\psi(\check{\pi}, \tilde{s})(r) \approx \min\{7.5(1 + 0.267\,L(r) - 0.267r), 16(-0.344 - 0.125\,L(r) + 0.125r)\} \vee 0$$

where $L(r) := \sqrt{(-7.396 + r)(-2.104 + r)}$, $r \in [0, 1]$. The graphs of $\tilde{s}$ and $\psi(\check{\pi}, \tilde{s})$ are given respectively in Fig. 2 and 3. Moreover, by analyzing the optimality equation (4.2), we can show that $\check{\pi} = (\pi_1, \pi_2, \pi_3, \ldots)$ with

$$\pi_t(\tilde{s}_t) = 1_{\{1\}} \; (t = 1, 2, \ldots)$$

is absolutely optimal.

Acknowledgments

The authors wish to thank the referees for suggestions which improved the paper.

References

[1] Ammar, E., and Metz, J., "On fuzzy convexity and parametric fuzzy optimization", *Fuzzy Sets and Systems* 49 (1992) 135–141.
[2] Baldwin, J.F. and Pilsworth, B.W., "Dynamic programming for fuzzy system with fuzzy environment," *Journal of Mathematical Analysis and Applications*, 85 (1982) 1–23.
[3] Bellman, R.E., and Zadeh, L.A., "Decision-making in a fuzzy environment," *Management Science Series B* 17 (1970) 141–164.
[4] Birkhoff, G., "Lattice theory", *American Mathematical Society Collected Publications* 25 (1940).
[5] Blackwell, D., "Discrete dynamic programming", *Annals of Mathematics and Statistics* 33 (1962) 719–726.
[6] Blackwell, D., "Discounted dynamic programming", *Annals of Methematics and Statistics* 36 (1965) 226–235.
[7] Xu, Chen-wei, "On convexity of fuzzy sets and fuzzy relations", *Information Science* 59 (1992) 92–102.
[8] Esogbue, A.O., and Bellman, R.E., "Fuzzy dynamic programming and its extensions", *TIMS/Studies in Management Science* 20 (1984) 147–167.
[9] Holzbaur, U.D., "A concept of fuzzy transition probabilities", *Annals of Operations Research* 32 (1991) 35–50.
[10] Kurano, M., Yasuda, M., Nakagami, J., and Yoshida, Y., "A limit theorem in some fuzzy dynamic systems", *Fuzzy Sets and Systems* 51 (1992) 83–88.
[11] Madan, L., Puri, L., and Ralescu, D.A., "The concept of normality for fuzzy random variables", *Annals of Probability* 13 (1985) 1373–1379.
[12] Nanda, S., "On sequences of fuzzy numbers", *Fuzzy Sets and Systems* 33 (1989) 123–126.
[13] Novák, V., *Fuzzy Sets and their Applications,* Adom Hilger, Bristol-Boston, 1989.
[14] Yoshida, Y., Yasuda, M., Nakagami, J., and Kurano, M., "A potential of fuzzy relations with a linear structure: The contractive case", *Fuzzy Sets and Systems* 60 (1993) 283–294.
[15] Yoshida, Y., Yasuda, M., Nakagami, J., and Kurano, M., "A potential of fuzzy relations with a linear structure: The unbounded case", *Fuzzy Sets and Systems* 66 (1994) 83–95.
[16] Zadeh, L.A., "Fuzzy sets", *Information and Control* 8 (1965) 338–353.