Markov decision processes with fuzzy rewards

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

Abstract

In this paper, we consider the model that the information on the rewards in vector-valued Markov decision processes includes imprecision or ambiguity. The fuzzy reward model is analyzed as follows: The fuzzy reward is represented by the fuzzy set on the multi-dimensional Euclidian space $\mathbb{R}^p$ and the infinite horizon fuzzy expected discounted reward(FEDR) from any stationary policy is characterized as a unique fixed point of the corresponding contractive operator. Also, we fined a Pareto optimal policy which maximizes the infinite horizon FEDR over all stationary policies under the pseudo order induced by a convex cone $\mathbb{R}^p$. As a numerical example, the machine maintenance problem is considered.

Keywords: Multi-dimensional fuzzy reward model, Markov decision process, Pareto optimal, fuzzy optimality equation.

AMS 1991 subject classification. Primary: 90c40; Secondary: 90c39.

1. Introduction

In mathematical modeling in terms of Markov decision processes (MDPs, in short, cf. [2, 6, 12, 15]), it often occurs that the information on the reward function includes imprecision or ambiguity. As an example, the reward earned in a day is about 700 dollars or closed to 700 dollars. On the other hand, multi-criteria decision making is typically involving flexible requirements for the optimality. In order to deal with uncertain data and flexible requirements we can use a fuzzy set representation (cf. [17]). In this paper, we consider the case that the $\mathbb{R}^p$-valued rewards in standard MDPs are specified by fuzzy sets on $\mathbb{R}^p$, where $\mathbb{R}^p$ is a p-dimensional Euclidean space ($p \geq 1$).

Recently, Kurano et al [10] has introduced a pseudo order $\preccurlyeq_K$ in the class of fuzzy sets on $\mathbb{R}^p$, which is a natural extension of fuzzy max order (cf.[5, 16]) in fuzzy numbers on $\mathbb{R}$ and induced by a convex cone K in $\mathbb{R}^p$. Under this pseudo order $\preccurlyeq_K$, we fined a Pareto optimal policy which maximizes the infinite horizon fuzzy expected discounted reward (FEDR) over all stationary policies. Associated with each stationary policy is a corresponding contractive operator on fuzzy sets, whose fixed point represents the infinite horizon FEDR. Moreover, the Pareto optimal policies are characterized by maximal solutions of an optimal equation including efficient fuzzy set functions. As a numerical example, the machine maintenance problem is considered.

For an interval or fuzzy treatment for MDPs with uncertain transition matrices, see [8, 9, 11] in which the intervals or fuzzy sets are used to describe uncertain transition matrices. Also, for the optimization of fuzzy dynamic system refer to [7, 19].

This paper is organized as follows: In Section 2, we shall give some notations needed for fuzzy treatments and a pseudo order relation of fuzzy sets on $\mathbb{R}^p$ is reviewed referring to Kurano et al [10] and the expectation of discrete fuzzy random variables is specified. In Section 3, we describe the fuzzy reward model and specify the optimization problem. In Section 4, the infinite horizon FEDR from a stationary policy is given as a fixed point of a corresponding operator, which is used to obtain the optimality equation and characterize a Pareto optimal policy in Section 5.

2. Preliminaries

We write fuzzy sets on $\mathbb{R}^p$ by their membership functions $\tilde{s}: \mathbb{R}^p \to [0,1]$ (see Novák [13] and Zadeh [20]). The α-cut ($\alpha \in [0,1]$) of the fuzzy set $\tilde{s}$ on $\mathbb{R}^p$ is defined as

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

where cl denotes the closure of the set. A fuzzy set $\tilde{s}$ is called convex if

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

where $a \wedge b = \min\{a, b\}$.

Note that $\tilde{s}$ is convex if and only if the α-cut $\tilde{s}_\alpha$ is a convex set for all $\alpha \in [0,1]$. Let $\mathcal{F}(\mathbb{R}^p)$ be the set of all convex fuzzy sets whose membership functions $\tilde{s}: \mathbb{R}^p \to [0,1]$ are upper-semicontinuous and normal ($\sup_{x \in \mathbb{R}^p} \tilde{s}(x) = 1$) and have a compact support. In the one-dimensional case $p = 1$, $\mathcal{F}(\mathbb{R})$ denotes the set of all fuzzy numbers. Let $\mathcal{C}(\mathbb{R}^p)$ be the set of all compact convex subsets of $\mathbb{R}^p$. We note that when $p = 1$, $\mathcal{F}(\mathbb{R})$ denotes the set of bounded and closed intervals in $\mathbb{R}$.

The definitions of addition and scalar multiplication on $\mathcal{F}(\mathbb{R}^p)$ are as follows: For $\tilde{s}, \tilde{r} \in \mathcal{F}(\mathbb{R}^p)$ and $\lambda \geq 0$,

$$(\tilde{s} + \tilde{r})(x) := \sup_{\substack{x_1, x_2 \in \mathbb{R}^p \\ x_1 + x_2 = x}} \{\tilde{s}(x_1) \wedge \tilde{r}(x_2)\}, \tag{2.1}$$

$$(\lambda \tilde{s})(x) := \begin{cases} \tilde{s}(x/\lambda) & \text{if } \lambda > 0 \\ 1_{\{0\}}(x) & \text{if } \lambda = 0 \end{cases} \quad (x \in \mathbb{R}^p), \tag{2.2}$$

where $1_{\{\cdot\}}(\cdot)$ is an indicator. By using set operations $A + B := \{x + y \mid x \in A, y \in B\}$ and $\lambda A := \{\lambda x \mid x \in A\}$ for any non-empty sets $A, B \subset \mathbb{R}^p$, the following holds immediately.

$$(\tilde{s} + \tilde{r})_\alpha = \tilde{s}_\alpha + \tilde{r}_\alpha \quad \text{and} \quad (\lambda \tilde{s})_\alpha = \lambda \tilde{s}_\alpha \quad (\alpha \in [0,1]). \tag{2.3}$$

Let ρ be the Hausdorff metric on $\mathcal{C}(\mathbb{R}^p)$, that is, for $A, B \in \mathcal{C}(\mathbb{R}^p)$,

$$\rho(A,B) = \max\{\max_{a\in A} d(a,B), \max_{b\in B} d(b,A)\},$$

where d is a metric in $\mathbb{R}^p$ and $d(x,D) = \min_{y\in D} d(x,y)$ for $x \in \mathbb{R}^p$ and $D \in \mathcal{C}(\mathbb{R}^p)$. Extending this ρ to $C(\mathbb{R}^p)$, we define, with abuse of notation, the Hausdorff metric on $\mathcal{F}(\mathbb{R}^p)$ by

$$\rho(\widetilde{u},\widetilde{v}) = \sup_{\alpha\in[0,1]} \rho(\widetilde{u}_\alpha, \widetilde{v}_\alpha) \quad \text{for } \widetilde{u},\widetilde{v} \in \mathcal{F}(\mathbb{R}^p), \tag{2.4}$$

where $\widetilde{u}_\alpha$ and $\widetilde{v}_\alpha$ are α-cuts of $\widetilde{u}$ and $\widetilde{v}$ respectively.

Then, the following facts are well known.

Lemma 2.1 (cf. [14]). *The metric space $(\mathcal{F}(\mathbb{R}^p), \rho)$ is complete.*

Lemma 2.2 (cf. [3]). *If $\widetilde{u}, \widetilde{v}, \widetilde{u'}, \widetilde{v'}$ and $\widetilde{r} \in \mathcal{F}(\mathbb{R}^p)$, then*

(i) $\rho(\lambda\widetilde{u}, \lambda\widetilde{v}) = \lambda\rho(\widetilde{u},\widetilde{v})$ *for all* $\lambda \geq 0$.

(ii) $\rho(\widetilde{u}+\widetilde{u'}, \widetilde{v}+\widetilde{v'}) \leq \rho(\widetilde{u},\widetilde{v}) + \rho(\widetilde{u'},\widetilde{v'})$,

(iii) $\rho(\widetilde{r}+\widetilde{u}, \widetilde{r}+\widetilde{v}) = \rho(\widetilde{u},\widetilde{v})$.

Here, we pick up a pseudo order relation introduced in Kurano et al [10], which is necessary for our problem formulation in the sequel. The partial order relation $\preccurlyeq_1$ on $\mathcal{C}(\mathbb{R})$ is defined as follows: For any $[c_1,c_2]$, $[c_1',c_2'] \in \mathcal{C}(\mathbb{R})$, $[c_1,c_2] \preccurlyeq_1 [c_1',c_2']$ means that $c_1 \leq c_1'$ and $c_2 \leq c_2'$.

Let K be a non-empty cone of $\mathbb{R}^p$. Using this K, we can define a pseudo order relation $\preccurlyeq_K$ on $\mathbb{R}^p$ by $x \preccurlyeq_K y$ if and only if $y - x \in K$. By the pseudo order $\preccurlyeq_K$ on $\mathbb{R}^p$, a pseudo order $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R}^p)$ is defined as follows.

For $\widetilde{s},\widetilde{r} \in \mathcal{F}(\mathbb{R}^p)$, $\widetilde{s} \preccurlyeq_K \widetilde{r}$ means the following (F.a) *and* (F.b)*:*

(F.a) *For any $x \in \mathbb{R}^p$, there exists $y \in \mathbb{R}^p$ such that $x \preccurlyeq_K y$ and $\widetilde{s}(x) \leq \widetilde{r}(y)$.*

(F.b) *For any $y \in \mathbb{R}^p$, there exists $x \in \mathbb{R}^p$ such that $x \preccurlyeq_K y$ and $\widetilde{s}(x) \geq \widetilde{r}(y)$.*

When $p = 1$ and $K = [0,\infty)$, the $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R})$ is a partial order and called the fuzzy max order (cf. [5, 16]) defined by $\preccurlyeq_1$. That is, for $\widetilde{s},\widetilde{r} \in \mathcal{F}(\mathbb{R})$, $\widetilde{s} \preccurlyeq_1 \widetilde{r}$ means that $\widetilde{s}_\alpha^L \leq \widetilde{r}_\alpha^L$ and $\widetilde{s}_\alpha^U \leq \widetilde{r}_\alpha^U$ for all $\alpha \in [0,1]$, where the α-cuts of $\widetilde{s}$ and $\widetilde{r}$ are denoted respectively by $\widetilde{s}_\alpha = [\widetilde{s}_\alpha^L, \widetilde{s}_\alpha^U]$ and $\widetilde{r}_\alpha = [\widetilde{r}_\alpha^L, \widetilde{r}_\alpha^U]$.

Define the dual cone of a cone K by

$$K^+ := \{a \in \mathbb{R}^p \mid a \cdot x \geq 0 \text{ for all } x \in K\},$$

where $x \cdot y$ denotes the inner product on $\mathbb{R}^p$ for $x, y \in \mathbb{R}^p$. For a subset $A \subset \mathbb{R}^p$ and $a \in \mathbb{R}^p$, we define

$$a \cdot A := \{a \cdot x \mid x \in A\} \ (\subset \mathbb{R}). \tag{2.5}$$

The definition (2.5) means the projection of A on the extended line of the vector a if $a \cdot a = 1$. It is trivial that $a \cdot A \in \mathcal{C}(\mathbb{R})$ if $A \in \mathcal{C}(\mathbb{R}^p)$ and $a \in \mathbb{R}^p$.

The pseudo order relation $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R}^p)$ is characterized by $\preccurlyeq_1$ on $\mathcal{F}(\mathbb{R})$ through the projection (2.5), where the proof is in [10].

Lemma 2.3 [10]. *Let $\widetilde{u}, \widetilde{v} \in \mathcal{F}(\mathbb{R}^p)$. Then, $\widetilde{u} \preccurlyeq_K \widetilde{v}$ on $\mathcal{F}(\mathbb{R}^p)$ if and only if $a \cdot \widetilde{u}_\alpha \preccurlyeq_1 a \cdot \widetilde{v}_\alpha$ on $\mathcal{F}(\mathbb{R})$ for all $a \in K^+$ and $\alpha \in [0,1]$.*

Lemma 2.4 [10]. *Let a sequence $\{\widetilde{u}_l\} \subset \mathcal{F}(\mathbb{R}^p)$ be such that $\widetilde{u}_1 \preccurlyeq_K \widetilde{u}_2 \preccurlyeq_K \cdots$, and $\widetilde{u} = \lim_{l\to\infty} \widetilde{u}_l \in \mathcal{F}(\mathbb{R}^p)$. Then, it holds that $\widetilde{u}_1 \preccurlyeq_K \widetilde{u}$.*

The following lemma is used in the sequel.

Lemma 2.5. *Let $A, B \in \mathcal{C}(\mathbb{R}^p)$ and $a \in \mathbb{R}^p$. Then, we have:*

(i) $a \cdot (A + B) = a \cdot A + a \cdot B$,

(ii) $a \cdot (\lambda A) = \lambda(a \cdot A)$ *for all* $\lambda \geq 0$.

Proof. For $A, B \in \mathcal{C}(\mathbb{R}^p)$, $a \cdot (A+B) \in \mathcal{C}(\mathbb{R})$, so that, $a \cdot (A+B) = [a \cdot (x^L + y^L), a \cdot (x^U + y^U)]$ for some $x^L, x^U \in A$ and $y^L, y^U \in B$. Since $a \cdot (x^L + y^L) = a \cdot x^L + a \cdot y^L$, $a \cdot x^L \in a \cdot A$ and $a \cdot y^L \in a \cdot B$, it holds $a \cdot (x^L + y^L) \in a \cdot A + a \cdot B$. Similarly, $a \cdot (x^U + y^U) \in a \cdot A + a \cdot B$. Thus, $a \cdot (A+B) \subset a \cdot A + a \cdot B$. Conversely, if we set $a \cdot A = [a \cdot x^L, a \cdot x^U]$ and $a \cdot B = [a \cdot y^L, a \cdot y^U]$, $a \cdot (A + B) = [a \cdot (x^L + y^L), a \cdot (x^U + y^U)]$, which implies $a \cdot A + a \cdot B \subset a \cdot (A + B)$.

Also, (ii) clearly holds, as required. □

In order to formulate the optimization problem in the next section, we need the concept of the expectation of discrete fuzzy random variables.

Let $(\Omega, \mathcal{B}, P)$ be a probability space and $\widetilde{X} : \Omega \to \mathcal{F}(\mathbb{R}^p)$ a discrete fuzzy random variable with its range $\{\widetilde{s}_1, \widetilde{s}_2, \cdots, \widetilde{s}_l\} \subset \mathcal{F}(\mathbb{R}^p)$. Then, we define the expectation of $\widetilde{X}$ by

$$E[\widetilde{X}] = \sum_{i=1}^{l} \widetilde{s}_i P(\widetilde{X} = \widetilde{s}_i). \tag{2.6}$$

Note that the expectation in (2.6) is defined in (2.1) and (2.2). The definition of (2.6) is corresponding to the discrete case of the integral of a set-valued function (cf. [1]) or the expectation of general fuzzy random variables (cf. [14]).

The following clearly holds.

Lemma 2.5. *If $\widetilde{X}$ and $\widetilde{Y}$ are discrete fuzzy random variables whose ranges are finite subsets of $\mathcal{F}(\mathbb{R}^p)$, then*

(i) $E[\widetilde{X}] \in \mathcal{F}(\mathbb{R}^p)$,

(ii) $E[\widetilde{X} + \widetilde{Y}] = E[\widetilde{X}] + E[\widetilde{Y}]$,

(iii) $E[\lambda \widetilde{X}] = \lambda E[\widetilde{X}]$ *for all* $\lambda \geq 0$.

3. The fuzzy reward model

In this section, we formulate MDPs with fuzzy rewards on $\mathbb{R}^p$ and specify our optimization problem. Let S and A be finite sets denoted by $S = \{1, 2, \cdots, n\}$ and $A = \{1, 2, \cdots, k\}$. The sequential decision model consists of four objects:

$$(S, A, \{q_{ij}(a); i, j \in S,\ a \in A\}, \tilde{r}),$$

where S and A denote the state and action spaces respectively and $\tilde{r} = \tilde{r}(i, a) \in \mathcal{F}(\mathbb{R}^p)$ is a fuzzy reward function on $S \times A$ and $\{q_{ij}(a)\}$ is the law of motion, i.e., for each $(i, a) \in S \times A$, $q_{ij}(a) \geq 0$ and $\sum_{j \in S} q_{ij}(a) = 1$.

When the system is in state $i \in S$ and we take action $a \in A$, the present state moves to a new state $j \in S$ selected according to the probability distribution $q_{i\cdot}(a)$ and we receive a fuzzy reward $\tilde{r}(i, a) \in \mathcal{F}(\mathbb{R}^p)$. This process is then repeated from the new state $j \in S$. The sample space is the product space $\Omega = (S \times A)^\infty$ such that the projection X_t and Δ_t on the t-th factor S and A describe the state and the action at the t-th time of the process $(t = 1, 2, \ldots)$.

We denote by F the set of all functions from S to A. A policy π is a sequence $(f_1, f_2, \ldots)$ of functions with $f_t \in F$ $(t \geq 1)$. Let Π be the class of policies. We denote by f^∞ the policy $(f_1, f_2, \ldots)$ with $f_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary and denoted simply by $f \in F$. The set of all stationary policies will be denoted by Π_F. Then, for each policy $\pi \in \Pi$ and starting state $i \in S$, we can define the probability measure P^i_π on Ω in a usual way. Here, we consider the expected fuzzy reward in which the future reward is discounted with a factor β $(0 < \beta < 1)$.

For any policy $\pi \in \Pi$ and starting state $i \in S$, let

$$\tilde{\phi}_T(i, \pi) = \sum_{t=1}^{T} \beta^{t-1} E^i_\pi[\tilde{r}(X_t, \Delta_t)], \tag{3.1}$$

where E^i_π is the expectation with respect to P^i_π and its expectation of fuzzy random variable is defined by (2.6). We note from Lemma 2.5 that $\tilde{\phi}_T(i, \pi) \in \mathcal{F}(\mathbb{R}^p)$ for $i \in S$, $\pi \in \Pi$ and $T \geq 1$.

In order to rewrite (3.1) by using vectors and matrices, we shall introduce some notations. Let $\mathcal{F}(\mathbb{R}^p)^n$ be the set of all n-dimensional column vectors whose elements are in $\mathcal{F}(\mathbb{R}^p)$, i.e.,

$$\mathcal{F}(\mathbb{R}^p)^n := \{\tilde{u} = (\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n)' \mid \tilde{u}_i \in \mathcal{F}(\mathbb{R}^p), 1 \leq i \leq n\},$$

where d' denotes the transpose of a vector d.

The Hausdorff metric ρ on $\mathcal{F}(\mathbb{R}^p)^n$ is defined (with abuse of notation) by

$$\rho(\tilde{u}, \tilde{v}) = \max_{1 \leq i \leq n} \rho(\tilde{u}_i, \tilde{v}_i),$$

where $\widetilde{u} = (\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n)'$, $\widetilde{v} = (\widetilde{v}_1, \widetilde{v}_2, \dots, \widetilde{v}_n)' \in \mathcal{F}(\mathbb{R}^p)^n$ and $\rho(\widetilde{u}_i, \widetilde{v}_i)$ is defined in (2.4). Then, from Lemma 2.1, we observe that the metric space $(\mathcal{F}(\mathbb{R}^p)^n, \rho)$ is complete.

For a $n \times n$ stochastic matrix $Q = (q_{ij})$ and $\widetilde{u} = (\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n)' \in \mathcal{F}(\mathbb{R}^p)^n$, the product $Q\widetilde{u} \in \mathcal{F}(\mathbb{R}^p)^n$ will be defined by

$$(3.2) \qquad (Q\widetilde{u})_i = \sum_{j=1}^{n} q_{ij}\widetilde{u}_j \quad (1 \le i \le n).$$

Here, we associate with each $f \in F$ the n-dimensional column fuzzy vector $\widetilde{r}(f) \in \mathcal{F}(\mathbb{R}^p)^n$ whose i-th element is $\widetilde{r}(i, f(i)) \in \mathcal{F}(\mathbb{R}^p)$ and the $n \times n$ stochastic matrix $Q(f)$ whose (i,j) element is $q_{ij}(f(i))$. For each policy $\pi \in \Pi$, let

$$\widetilde{\phi}_T(\pi) = (\widetilde{\phi}_T(1,\pi), \widetilde{\phi}_T(2,\pi), \dots, \widetilde{\phi}_T(n,\pi)) \in \mathcal{F}(\mathbb{R}^p)^n \quad (T \ge 1).$$

Then, we have the following.

Lemma 3.1. *For any $\pi = (f_1, f_2, \dots) \in \Pi$, we have:*

(i) *$\widetilde{\phi}_T(\pi)$ is described by the following matrix representation.*

$$(3.3) \qquad \widetilde{\phi}_T(\pi) = \widetilde{r}(f_1) + \beta Q_1(\pi)\widetilde{r}(f_2) + \cdots + \beta^{T-1}Q_{T-1}(\pi)\widetilde{r}(f_T) \quad (T \ge 1),$$
$$\text{where } Q_t(\pi) = Q(f_1)\cdots Q(f_t) \quad (t \ge 1).$$

(ii) *$\{\widetilde{\phi}_T(\pi)\}_{T=1}^{\infty}$ is a Cauchy sequence.*

Proof. By the definition, for any $t \ge 0$ we have that

$$E_\pi^i = \sum_{j \in S} P_\pi^i(X_t = j)\widetilde{r}(j, f_t(j)) = \sum_{j \in S} Q_t(\pi)_{ij}\widetilde{r}(j, f_t(j)),$$

which clearly leads to (3.2).

For any $T > H$, it holds from Lemma 2.2 that

$$\rho(\widetilde{\phi}_T(\pi), \widetilde{\phi}_H(\pi)) \le \rho(\widetilde{0}, \sum_{t=H+1}^{T} \beta^{t-1}Q_{t-1}\widetilde{r}(f_t))$$
$$= \beta^H \rho(\widetilde{0}, \sum_{t=H+1}^{T} \beta^{t-H-1}Q_t\widetilde{r}(f_t)) \le \beta^H \rho(\widetilde{0}, \widetilde{r})/(1-\beta),$$

where $\widetilde{0} \equiv \mathbf{1}_{\{0\}}$. This implies (ii), as required. □

By Lemma 3.1, the infinite horizon FEDR from π can be defined by

$$\widetilde{\phi}(\pi) = \lim_{T \to \infty} \widetilde{\phi}_T(\pi).$$

In order to specify our optimization problem, we extend the pseudo order $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R}^p)$ given in the preceding section to that on $\mathcal{F}(\mathbb{R}^p)^n$ as follows: For $\widetilde{u} = (\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n)'$, $\widetilde{v} = (\widetilde{v}_1, \widetilde{v}_2, \dots, \widetilde{v}_n)' \in \mathcal{F}(\mathbb{R}^p)^n$, $\widetilde{u} \preccurlyeq_K \widetilde{v}$ means $\widetilde{u}_i \preccurlyeq_K \widetilde{v}_i$ for all i $(1 \le i \le n)$.

Then, our problem is to maximize the $\widetilde{\phi}(\pi) \in \mathcal{F}(\mathbb{R}^p)^n$ over all policies $\pi \in \Pi$ with respect to the pseudo order $\preccurlyeq_K$.

4. Stationary policies and operators

In this section, the infinite horizon FEDR from a stationary policy is given as a unique fixed point of a corresponding operator. Associated with each $f \in F$ is a corresponding operator $U_f : \mathcal{F}(\mathbb{R}^p)^n \to \mathcal{F}(\mathbb{R}^p)^n$ defined as follows: For $\widetilde{u} \in \mathcal{F}(\mathbb{R}^p)^n$,

$$U_f\widetilde{u} = \widetilde{r}(f) + \beta Q(f)\widetilde{u}, \tag{4.1}$$

where the arithmetics in (4.1) are defined in the preceding sections.

Since it holds that $\lambda(\widetilde{c} + \widetilde{d}) = \lambda\widetilde{c} + \lambda\widetilde{d}$ for any $\widetilde{c}, \widetilde{d} \in \mathcal{F}(\mathbb{R}^p)$ and $\lambda \geq 0$, the following lemma is easily proved.

Lemma 4.1. *If Q is $n \times n$ stochastic matrix and $\widetilde{u}, \widetilde{v} \in \mathcal{F}(\mathbb{R}^p)^n$, then it holds that*

$$Q(\widetilde{u} + \widetilde{v}) = Q\widetilde{u} + Q\widetilde{v}.$$

For any policy $\pi = (f_1, f_2, \dots)$, let $\pi^{-l} = (f_{l+1}, f_{l+2}, \dots)$ for each $l \geq 1$. The sequence $\{\widetilde{\phi}_T(\pi)\}_{T=1}^{\infty}$ is recursively described.

Lemma 4.2. *For any policy $\pi = (f_1, f_2, \dots)$, we have*

$$\widetilde{\phi}_T(\pi) = U_{f_1}U_{f_2}\cdots U_{f_l}\widetilde{\phi}_{T-l}(\pi^{-l}) \quad \text{for each } \; l \geq 1. \tag{4.2}$$

Proof. Since $\widetilde{\phi}_1(\pi^{-1}) = \widetilde{r}(f_2)$, we have

$$\widetilde{\phi}_2(\pi) = \widetilde{r}(f_1) + \beta Q(f_1)\widetilde{r}(f_2) = U_{f_1}\widetilde{\phi}_1(\pi).$$

For $T = 3$, from Lemma 4.1, we have that

$$\widetilde{\phi}_3(\pi) = \widetilde{r}(f_1) + \beta Q(f_1)\widetilde{r}(f_2) + \beta^2 Q(f_1)Q(f_2)\widetilde{r}(f_3)$$

$$= \widetilde{r}(f_1) + \beta Q(f_1)\big(\widetilde{r}(f_2) + \beta Q(f_2)\widetilde{r}(f_3)\big) = U_{f_1}\widetilde{\phi}_2(\pi^{-1}).$$

By induction on T and l, we can easily prove (4.2). □

Here are some basic properties of U_f. The following lemma is easily proved from Lemma 2.2.

Lemma 4.3. *For $f \in F$, U_f is a contraction with modulus β, i.e.,*

$$\rho(U_f\widetilde{u}, U_f\widetilde{v}) \leq \beta\rho(\widetilde{u}, \widetilde{v}), \quad \text{for } \; \widetilde{u}, \widetilde{v} \in \mathcal{F}(\mathbb{R}^p)^n.$$

Lemma 4.4. *Let K be a convex cone of $\mathbb{R}^p$. Then, for $f \in F$, U_f is monotone with respect to the pseudo order $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R}^p)^n$, i.e., for any $\widetilde{u}, \widetilde{v} \in \mathcal{F}(\mathbb{R}^p)^n$ with $\widetilde{u} \preccurlyeq_K \widetilde{v}$, it holds that $U_f\widetilde{u} \preccurlyeq_K U_f\widetilde{v}$.*

Proof. From Lemma 2.3, it suffices to show that $a \cdot (U_f\widetilde{u})_{i,\alpha} \preccurlyeq_1 a \cdot (U_f\widetilde{v})_{i,\alpha}$ for all $a \in K$, $\alpha \in [0,1]$ and $i = 1, 2, \ldots, n$, where $(U_f\widetilde{v})_{i,\alpha}$ is the α-cut of the i-th element of $U_f\widetilde{v}$.

Applying Lemma 2.5, we get

$$a \cdot (U_f\widetilde{v})_{i,\alpha} = a \cdot \widetilde{r}(i, f(i))_\alpha + \beta \sum_{j=1}^{n} q_{ij}(f(i))(a \cdot \widetilde{u}_{j,\alpha}). \tag{4.2}$$

Since $\widetilde{u} \preccurlyeq_K \widetilde{v}$ implies from Lemma 2.3 that $a \cdot \widetilde{u}_{j,\alpha} \preccurlyeq_1 a \cdot \widetilde{v}_{j,\alpha}$ for all $j = 1, 2, \ldots, n$, (4.2) implies that $a \cdot (U_f\widetilde{u})_{j,\alpha} \preccurlyeq_1 a \cdot (U_f\widetilde{v})_{j,\alpha}$. This completes the proof. □

By Lemma 4.2, $\widetilde{\phi}_T(f) = U_f\widetilde{\phi}_{T-1}(f)$ for all $T \geq 2$. As $T \to \infty$ in the above, $\widetilde{\phi}(f)$ is a fixed point of U_f. Thus, noting Lemma 4.3, the characterization of $\widetilde{\phi}(f)$ is immediately formulated as a theorem.

Theorem 4.1. *For any stationary policy $f \in F$, $\widetilde{\phi}(f)$ is a unique solution of the following equation:*

$$\widetilde{u} = U_f\widetilde{u}, \quad \widetilde{u} \in \mathcal{F}(\mathbb{R}^p)^n. \tag{4.3}$$

Note that (4.3) can be rewritten as the α-cut equation:

$$\widetilde{u}_\alpha = \widetilde{r}(f)_\alpha + \beta Q(f)\widetilde{u}_\alpha, \quad \alpha \in [0,1], \tag{4.4}$$

where $\widetilde{u}_\alpha = (\widetilde{u}_{1,\alpha}, \widetilde{u}_{2,\alpha}, \ldots, \widetilde{u}_{n,\alpha})'$ and $\widetilde{r}(f)_\alpha = (\widetilde{r}(1, f(1))_\alpha, \widetilde{r}(2, f(2))_\alpha, \ldots, \widetilde{r}(n, f(n))_\alpha)' \in \mathcal{C}(\mathbb{R}^p)^n$.

From a contraction of U_f, the next corollary holds.

Corollary 4.1. *For any stationary policy $f \in F$,*

$$\widetilde{\phi}(f) = \lim_{l \to \infty} U_f^l \widetilde{u} \quad (\widetilde{u} \in \mathcal{F}(\mathbb{R}^p)^n).$$

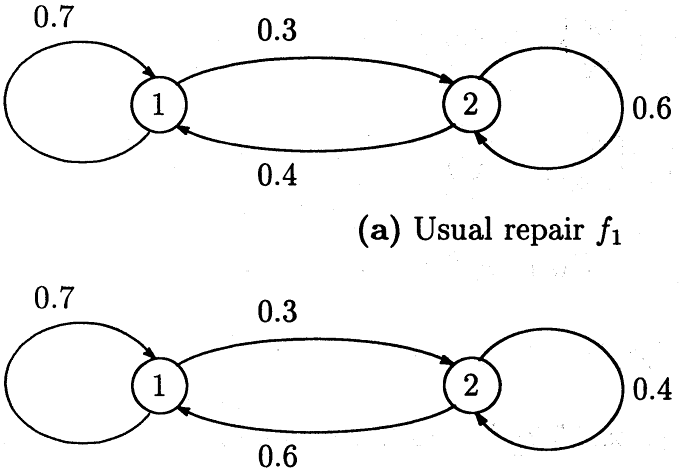
As a simple example, we consider a fuzzy treatment for a machine maintenance problem dealt with in ([12], p.1, p.17-18).

a machine maintenance problem. A machine can be operated synchronously, say, once an hour. At each period there are two states; one is operating(state 1), and the other is in failure(state 2). If the machine fails, it can be restored to perfect functioning by repair. At each period, if the machine is running, we earn the fuzzy return of (2,3,4) dollars per period; the probability of being in state 1 at the next step is 0.7 and the probability of moving to state 2 is 0.3 where for any $a < b < c$, the fuzzy number (a, b, c) on $\mathbb{R}$ is defined by

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

If the machine is in failure, we have two actions to repair the failed machine; one is a usual repair, denoted by 1, that yields the fuzzy reward $(-2,-1,0)$ dollars with the probability 0.4 moving in state 1 and the probability 0.6 being in state 2; another is a rapid repair, denoted by 2, that requires the fuzzy reward $(-3,-2,-1)$ dollars with the probability 0.6 moving in state 1 and the probability 0.4 being in state 2.

For the model considered, $S=\{1,2\}$ and there exists two stationary policies, $F=\{f_1,f_2\}$ with $f_1(2)=1$ and $f_2(2)=2$, where f_1 denotes a policy of the usual repair and f_2 a policy of the rapid repair. The state transition diagrams and fuzzy reward vector for two policies are shown in Figure 1.



$$\tilde{r}(f_1)=\begin{pmatrix}(\ 2,\ 3,\ 4)\\(-2,-1,\ 0)\end{pmatrix}$$

(a) Usual repair f_1

$$\tilde{r}(f_2)=\begin{pmatrix}(\ 2,\ 3,\ 4)\\(-3,-2,-1)\end{pmatrix}$$

(b) Rapid repair f_2

Figure.1 Transition diagrams and fuzzy rewards.

Applying Theorem 4.1, we obtain the infinite horizon FEDR as a unique solution of (4.3). So, putting

$$\tilde{\phi}(f_1)_\alpha=([x_\alpha^1,y_\alpha^1],[x_\alpha^2,y_\alpha^2])',$$

the α-cut interval equations (4.4) with $\beta=0.9$ become:

$$\begin{aligned}x_\alpha^1&=2+\alpha+0.9(0.7x_\alpha^1+0.3x_\alpha^2)\\y_\alpha^1&=4-\alpha+0.9(0.7y_\alpha^1+0.3y_\alpha^2)\\x_\alpha^2&=-2+\alpha+0.9(0.4x_\alpha^1+0.6x_\alpha^2)\\y_\alpha^2&=-\alpha+0.9(0.4y_\alpha^1+0.6y_\alpha^2)\end{aligned}$$

After a simple calculation, we obtain

$$\tilde{\phi}(f_1)_\alpha=\Big([10\alpha+\frac{380}{73},\frac{1840}{73}-10\alpha],[10\alpha-\frac{20}{73},\frac{1440}{73}-10\alpha]\Big)',$$

which leads to

$$\tilde{\phi}(f_1)=\Big((\frac{380}{73},\frac{1110}{73},\frac{1840}{73}),(-\frac{20}{73},\frac{710}{73},\frac{1440}{73})\Big)'.$$

5. Pareto optimality

Here, we confine our attention to the class of stationary policies, which simplifies our discussion in the sequel. Let K be a convex cone in $\mathbb{R}^p$. A policy $f^* \in \Pi_F$ is called Pareto optimal if there does not exist $f \in \Pi_F$ such that $\widetilde{\phi}(f^*) \preccurlyeq_K \widetilde{\phi}(f)$. In this section, we derive the optimal equation, by which Pareto optimal policies are characterized.

The following important result is crucial to the development in the characterization of Pareto optimality.

Lemma 5.1. *For any $f, g \in F$, suppose that*

$$\widetilde{\phi}(f) \begin{Bmatrix} \preccurlyeq_K \\ \prec_K \end{Bmatrix} U_g\widetilde{\phi}(f). \tag{5.1}$$

Then, it holds that

$$\widetilde{\phi}(f) \begin{Bmatrix} \preccurlyeq_K \\ \prec_K \end{Bmatrix} \widetilde{\phi}(g). \tag{5.2}$$

Proof. Suppose that $\widetilde{\phi}(f)$ $\{\substack{\preccurlyeq_K \\ \prec_K}\}$ $U_g\widetilde{\phi}(f)$. Then, we have from Lemma 4.3 that

$$\widetilde{\phi}(f) \quad \begin{Bmatrix} \preccurlyeq_K \\ \prec_K \end{Bmatrix} \quad U_g\widetilde{\phi}(f) \quad \preccurlyeq \quad U_g^l\widetilde{\psi}(f) \quad (l \geq 2),$$

So, taking the limit in the above as $l \to \infty$, (5.2) follows from Corollary 4.1 and Lemma 2.4. □

Let D be an arbitrary subset of $\mathcal{F}(\mathbb{R}^p)^n$. A point $\widetilde{u} \in D$ is called an efficient element of D with respect to $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R}^p)^n$ if and only if it holds that there does not exist $\widetilde{v} \in D$ such that $\widetilde{u} \prec_K \widetilde{v}$. We denote by eff$(D)$ the set of all elements of D efficient with respect to $\preccurlyeq_K$ on $\mathcal{F}(\mathbb{R}^p)^n$. For any $\widetilde{u} \in \mathcal{F}(\mathbb{R}^p)^n$, let $\mathcal{U}(\widetilde{u}) :=$ eff$(\{U_f\widetilde{u} \mid f \in F\})$. Note that $\mathcal{U}(\widetilde{u}) \subset \mathcal{F}(\mathbb{R}^p)^n$.

Here, we consider the following fuzzy equation including efficient fuzzy functions $\mathcal{U}(\cdot)$ on $\mathcal{F}(\mathbb{R}^p)^n$:

$$\widetilde{u} \in \mathcal{U}(\widetilde{u}), \quad \widetilde{u} \in \mathcal{F}(\mathbb{R}^p)^n. \tag{5.3}$$

The equation (5.3) is called an optimality equation, by which Pareto optimal policies are characterized. A solution $\widetilde{u}$ of (5.3) is called maximal if there does not exist any solution $\widetilde{u}'$ of (5.3) such that $\widetilde{u} \prec_K \widetilde{u}'$. Pareto optimal policies are characterized by maximal solutions of the optimality equation (5.3).

Theorem 5.1. *A policy f is Pareto optimal if and only if a fixed point of the corresponding U_f, $\widetilde{\phi}(f)$, is a maximal solution to the optimal equation (5.3).*

Proof. The proof of "only if "part is easily obtained from Lemma 5.1. In order to prove "if "part, suppose that $\widetilde{\phi}(f)$ is a maximal solution of (5.3) but f is not Pareto optimal. Then, there exists $f^{(1)} \in F$ such that $\widetilde{\phi}(f) \prec_K \widetilde{\phi}(f^{(1)})$.

Now, suppose that $\tilde{\phi}(f^{(1)}) \notin \mathrm{eff}(\tilde{\phi}(f^{(1)}))$. This assumption assures that there exists $f^{(2)} \in F$ satisfying $\tilde{\phi}(f^{(1)}) \prec_K U_{f^{(2)}}\tilde{\phi}(f^{(1)})$, which implies from (5.1) that $\tilde{\phi}(f^{(1)}) \prec_K \tilde{\phi}(f^{(2)})$. By repeating this method successively, we come to the conclusion that there exists $f^{(l)} \in F$ such that $\tilde{\phi}(f) \prec_K \tilde{\phi}(f^{(l)})$ and $\tilde{\phi}(f^{(l)})$ satisfies (5.3), which contradicts that $\tilde{\phi}(f)$ is maximal, as required. □

Remark. For vector-valued discounted MDPs, Furukawa[4] and White[18] had derived the optimality equation including efficient set-function on $\mathbb{R}^p$, by that Pareto optimal policies are characterized. The form of the optimal equation (5.3) is corresponding to a fuzzy version of MDPs.

For the machine maintenance problem given in Section 4, we find that

$$U_{f_2}\tilde{\phi}(f_1) = \Big((\frac{380}{73}, \frac{1110}{73}, \frac{1840}{73}), (-\frac{21}{73}, \frac{709}{73}, \frac{1439}{73}) \Big)',$$

Recall that

$$U_{f_1}\tilde{\phi}(f_1) = \tilde{\phi}(f_1) = \Big((\frac{380}{73}, \frac{1110}{73}, \frac{1840}{73}), (-\frac{20}{73}, \frac{710}{73}, \frac{1440}{73}) \Big)',$$

which satisfies $U_{f_2}\tilde{\phi}(f_1) \prec_1 \tilde{\phi}(f_1)$, where $\prec_1$ is the fuzzy max order on $\mathcal{F}(\mathbb{R})^2$ and corresponding to $\preccurlyeq_K$ in case of $K = [0, \infty)$.

Thus, $\tilde{\phi}(f_1) \in \mathrm{eff}(\{U_f\tilde{\phi}(f_1) \mid f \in F)$, so that from Theorem 5.1 f_1 is Pareto optimal. In fact, we can find, by solving (4.3) or (4.4) for f_2, that

$$\tilde{\phi}(f_2) = \Big((\frac{470}{91}, \frac{1380}{91}, \frac{2290}{91}), (-\frac{30}{91}, \frac{880}{91}, \frac{1790}{91}) \Big)', \text{ and } \tilde{\phi}(f_2) \prec_1 \tilde{\phi}(f_1).$$

References

[1] Aumann, R. J., Integrals of set-valued functions, *J. Math. Anal. Appl.* **12** (1965), 1-12.

[2] Blackwell, D., Discrete dynamic programming, *Ann. Math. Stat.* **33** (1962), 719-726.

[3] Diamond, P. and Kloeden, P., *Metric Spaces of Fuzzy Sets, Theory and Applications,* (1994), World Scientific.

[4] Furukawa, N., Characterization of optimal policies in vector-valued Markovian decision processes, *Math. Oper. Res.* **5** (1980), 271-279.

[5] Furukawa, N., Parametric orders on fuzzy numbers and their roles in fuzzy optimization problems, *Optimization* **40** (1997), 171-192.

[6] Howard, R., *Dynamic Programming and Markov processes,* (1960), MIT Press, Cambridge MA.

[7] Kurano, M., Yasuda, M., Nakagami, J. and Yoshida, Y., Markov-type fuzzy decision processes with a discounted reward on a closed interval, *European J. Oper. Res.*, **92** (1996), 649-662.

[8] Kurano, M., Song, J., Hosaka, M. and Huang, Y., Controlled Markov Set-Chains with Discounting, *J. Appl. Prob.*, **35** (1998), 293-302.

[9] Kurano, M., Yasuda, M. and Nakagami, J., Interval methods for uncertain Markov decision processes, *submitted to the International Workshop on Markov Processes and Controlled Markov Chains*, Changsha, Husan, China on August 22-28, 1999.

[10] Kurano, M., Yasuda, M., Nakagami, J. and Yoshida, Y., Ordering of fuzzy sets – A brief survey and new results, *J. Operations Research Society of Japan*, **43** (2000), 138-148.

[11] Kurano, M., Yasuda, M., Nakagami, J. and Yoshida, Y., A fuzzy treatment of uncertain Markov decision processes, Proceedings of ASSM2000 International Conference on Applied Stochastic System Modeling (Kyoto, March 2000), 148-157.

[12] Mine, H. and Osaki, S., *Markov Decision Processes*, (1970), Elsevier, Amsterdam.

[13] Novák, V., *Fuzzy sets and their applications*, (1989), Adam Hilger, Bristole-Boston.

[14] Puri, M. L. and Ralesca, D. A., Fuzzy random variable, *J. Math. Anal. Appl.*, **114** (1986), 402-422.

[15] Puterman, M. L., *Markov decision processes: Discrete Stochastic Dynamic Programming*, (1994), John Wiley & Sons, INC.

[16] J.Ramík and J.Řimánek, Inequality relation between fuzzy numbers and its use in fuzzy optimization, *Fuzzy Sets and Systems*, **16** (1985), 123-138.

[17] Stowínski, R., (ed.), *Fuzzy Sets in Decision Analysis, Operation Research and Statistics*, (1998), Kluwer Academic Publishers.

[18] White, D. J., Multi-objective infinite-horizon discounted Markov Decision Processes, *J. Math. Anal. Appl.*, **89** (1982), 639-647.

[19] Yoshida, Y., A time-average fuzzy reward criterion in fuzzy decision processes, *Information Sciences*, **110** (1998), 103-112.

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