KYBERNETIKA — VOLUME *58* (2022), NUMBER 2, PAGES 180–199

# MARKOV DECISION PROCESSES ON FINITE SPACES WITH FUZZY TOTAL REWARDS

Karla Carrero-Vera, Hugo Cruz-Suárez and Raúl Montes-de-Oca

The paper concerns Markov decision processes (MDPs) with both the state and the decision spaces being finite and with the total reward as the objective function. For such a kind of MDPs, the authors assume that the reward function is of a fuzzy type. Specifically, this fuzzy reward function is of a suitable trapezoidal shape which is a function of a standard non-fuzzy reward. The fuzzy control problem consists of determining a control policy that maximizes the fuzzy expected total reward, where the maximization is made with respect to the partial order on the $\alpha$-cuts of fuzzy numbers. The optimal policy and the optimal value function for the fuzzy optimal control problem are characterized by means of the dynamic programming equation of the standard optimal control problem and, as main conclusions, it is obtained that the optimal policy of the standard problem and the fuzzy one coincide and the fuzzy optimal value function is of a convenient trapezoidal form. As illustrations, fuzzy extensions of an optimal stopping problem and of a red-black gambling model are presented.

*Keywords:* Markov decision process, total reward, fuzzy reward, trapezoidal fuzzy number, optimal stopping problem, gambling model

*Classification:* 90C40, 93C40

## 1. INTRODUCTION

A common situation which appears in applied mathematics is that the data required to propose a mathematical model present ambiguity, vagueness or approximate characteristics of the problem under study, and a possibility to deal with this situation is to use the fuzzy theory proposed by L. Zadeh in his seminal article: [22]. The fuzzy theory is well established and there are extensions of this theory in several fields of mathematical sciences, such as control theory (see for instance [4] and [8]), and also it has impacted in applied areas (see for instance [9] and [10]).

In the fuzzy theory, the effectiveness of fuzzy number calculations lies in the form of the membership function (see [21]). Fuzzy numbers with a simpler membership function allow a more natural interpretation to be associated with it. A membership function with this characteristic is the one associated with trapezoidal fuzzy numbers (see [1, 6, 17], and [21]). Furthermore, the class of trapezoidal numbers contains the triangular fuzzy numbers which have been extensively studied and applied in fuzzy systems [14]. In

DOI: 10.14736/kyb-2022-2-0180

particular, the triangular fuzzy numbers could be used to approximate an arbitrary fuzzy number (see [2] and [23]).

Now, in this article, taking into account the previous considerations given on the fuzzy theory, the authors propose a Markov decision process (MDP, in plural MDPs) (see [16]) with finite state and decision spaces and fuzzy characteristics in its payoff or reward function. The main idea is the following: a standard model formulation of an MDP is considered (see Chapter 2 in [16]), with a nonnegative reward $R$ as a basis, and a new model is induced changing only $R$ by a reward function $\widetilde{R}$ with fuzzy values which is a function of $R$. Specifically, it is assumed that the fuzzy reward function is trapezoidal in shape. A relevant point is that a special case of $\widetilde{R}$ can be reduced to $R$ (in fact, $R$ can be trivially seen as a fuzzy number); hence, in this sense, an extension of the MDPs on finite spaces, nonnegative reward and total reward criterion theory is obtained. Also, this paper will refer to two kinds of MDPs: an MDP with a reward function $R$ which will be named the standard MDP or the standard model, and a fuzzy MDP or a fuzzy model which is an MDP with reward $\widetilde{R}$. Note that, in this work, the model formulation of the standard MDP has identical components of the fuzzy one, the only difference resides in the reward function; in particular, the set of all admissible policies is the same for both models.

Thus, with the information of the previous paragraph, the fuzzy control problem consists of determining a control policy that maximizes the expected total fuzzy reward, where the maximization is made with respect to the partial order on the $\alpha$-cuts of fuzzy numbers (see [11]). And, it is important to mention that the fuzzy optimal control problem is reduced to the standard optimal problem when $\widetilde{R}$ is reduced to $R$ seen as a fuzzy number (see Remark 5.4, below).

Moreover, the optimal policy and the optimal value function for the fuzzy optimal control problem are characterized by means of the dynamic programming equation of the standard optimal control problem and, as main consequences, it is obtained that the optimal policy of the standard problem and the fuzzy one coincide and the fuzzy optimal value function is of a convenient trapezoidal form.

To illustrate the theory developed, a fuzzy version of an optimal stopping problem (see [16]) and a fuzzy extension of the classical gambling model related to the ruin of a player are presented (see [18] and [19]).

Research works related to the topic developed here are the following: [12] and [20]. In both papers, versions of the discounted fuzzy control problem with finite state and action spaces are examined.

The paper is organized as follows. Section 2 presents the basic results about fuzzy numbers. The following section gives the standard theory on MDPs with the total reward criterion. Sections 4 and 5 present the theory on the fuzzy control problem under the criterion of the total reward. In Section 6 an optimal stopping problem and a gambling model to exemplify the theory developed are given and finally, in Section 7 the conclusions are provided.

**Notation.** In the article, the following standard mathematical symbols will be distinguished in the fuzzy context with an asterix symbol: " * " . That is, in the fuzzy context, "$\leq$", "$+$" and "$\sum$", will be denoted by "$\leq^*$", "$+^*$" and "$\sum^*$", respectively. Similarly,

in the fuzzy context, the expectation operator "$E$", the limit "lim" and the supremum "sup", will be denoted by "$E^*$", "$\lim^*$" and "$\sup^*$", respectively. It is important to mention that the product of a real number $\lambda$ and a fuzzy number $\Upsilon$ will be simply denoted as $\lambda\Upsilon$. Moreover, some special functions which appear as fuzzy quantities say, the reward function, the optimal value function, and so on, will be distinguished with a "tilde"; for instance, the fuzzy reward function will be written as $\widetilde{R}$.

## 2. PRELIMINARIES ON FUZZY THEORY

The first part of this section presents some definitions and basic results about the fuzzy set theory (see [7, 21], and [22]).

Let $\Lambda$ be a non-empty set. Then a *fuzzy set* $\Gamma$ on $\Lambda$ is defined in terms of the *membership function* $\Gamma'$, which assigns to each element of $\Lambda$ a real value from the interval $[0, 1]$. The $\alpha$-*cut* of $\Gamma$, denoted by $\Gamma_\alpha$, is defined to be the set $\Gamma_\alpha := \{x \in \Lambda \mid \Gamma'(x) \geq \alpha\}$ $(0 < \alpha \leq 1)$ and $\Gamma_0$ is the *closure* of $\{x \in \Lambda \mid \Gamma'(x) > 0\}$ denoted by $cl\{x \in \Lambda \mid \Gamma'(x) > 0\}$.

**Definition 2.1.** A *fuzzy number* $\Gamma$ is a fuzzy set defined on the set of real numbers $\mathbb{R}$ (i. e., taking $\Lambda = \mathbb{R}$ in the previous definition), which satisfies:

a) $\Gamma'$ is normal, i. e., there exists $x_0 \in \mathbb{R}$ with $\Gamma'(x_0) = 1$;

b) $\Gamma'$ is convex, i. e. $\Gamma_\alpha$ is convex for all $\alpha \in [0, 1]$;

c) $\Gamma'$ is upper-semicontinuous;

d) $\Gamma_0$ is compact.

The set of the fuzzy numbers will be denoted by $\mathfrak{F}(\mathbb{R})$.

**Definition 2.2.** A fuzzy number $\Gamma$ is called a *trapezoidal fuzzy number* if its membership function has the following form:

$$\Gamma'(x) = \begin{cases} 0 & \text{if } x \leq l \\ \frac{x-l}{m-l} & \text{if } l < x \leq m \\ 1 & \text{if } m < x \leq n \\ \frac{p-x}{p-n} & \text{if } n < x \leq p \\ 0 & \text{if } p < x\ , \end{cases} \tag{1}$$

where $l$, $m$, $n$ and $p$ are real numbers, with $l < m \leq n < p$. A trapezoidal fuzzy number is simply denoted by $(l, m, n, p)$.

**Remark 2.3.** a) The case in which $m = n$ in (1) will be named a *triangular fuzzy number* and it will be simply denoted as $(l, m, p)$. And, considering the degenerated case in which $l = m = p$ in a triangular number, the *fuzzy representation* of the real number $m$ is obtained with the membership function given by:

$$m'(x) = \begin{cases} 1 & \text{if } x = m \\ 0 & \text{if } x \neq m. \end{cases}$$

b) For a trapezoidal fuzzy number $\Gamma = (l, m, n, p)$ the corresponding $\alpha$-cuts are given by $\Gamma_\alpha = [(m-l)\alpha + l, p - (p-n)\alpha]$, $\alpha \in [0,1]$ (see [17]).

**Definition 2.4.** Let $\Gamma$ and $\Upsilon$ be fuzzy numbers. If " $\star$ " denotes the addition or the scalar mutiplication, then it is defined as a fuzzy set on $\mathbb{R}$, $\Gamma \star \Upsilon$, by the membership function:

$$(\Gamma \star \Upsilon)'(u) = \sup_{u=x\star y} \min\{\Gamma'(x), \Upsilon'(y)\},$$

for all $u \in \mathbb{R}$.

As a consequence of Definition 2.4, it is possible to obtain the following result for trapezoidal fuzzy numbers.

**Lemma 2.5.** (Rezvani and Molani [17]) If $H = (a_l, a_m, a_n, a_p)$ and $I = (b_l, b_m, b_n, b_p)$ be two trapezoidal fuzzy numbers and letting $\lambda$ be a positive number, then it follows that

a) $\lambda H = (\lambda a_l, \lambda a_m, \lambda a_n, \lambda a_p)$, and

b) $H +^* I = (a_l + b_l, a_m + b_m, a_n + b_n, a_p + b_p)$.

Let $\mathbb{D}$ denote the set of all closed bounded intervals on the real line $\mathbb{R}$. For $\Psi = [a_l, a_u]$, $\Phi = [b_l, b_u] \in \mathbb{D}$ define

$$d(\Psi, \Phi) = \max(|a_l - b_l|, |a_u - b_u|).$$

It is possible to verify that $d$ defines a metric on $\mathbb{D}$ and that $(\mathbb{D}, d)$ is a complete metric space (see [15]). Now, if $\widetilde{\eta} \in \mathfrak{F}(\mathbb{R})$, then, as its membership function satisfies b), c) and d) of Definition 2.1, it follows that $\eta_\alpha \in \mathbb{D}$. Therefore, it is defined that $\hat{d} : \mathfrak{F}(\mathbb{R}) \times \mathfrak{F}(\mathbb{R}) \longrightarrow \mathbb{R}$ by

$$\hat{d}(\widetilde{\eta}, \widetilde{\mu}) = \sup_{\alpha\in[0,1]} d(\eta_\alpha, \mu_\alpha). \tag{2}$$

It is straightforward to see that $\hat{d}$ is a metric in $\mathfrak{F}(\mathbb{R})$ (see [15]).

**Definition 2.6.** A sequence $\{\widetilde{\eta}_t\}_{t=0}^{\infty}$ of fuzzy numbers is said to be *convergent* to the fuzzy number $\widetilde{\mu}$, written as $\lim^*_{t\longrightarrow\infty} \widetilde{\eta}_t = \widetilde{\mu}$ if and only if $\hat{d}(\widetilde{\eta}_t, \widetilde{\mu}) \longrightarrow 0$ as $t \longrightarrow \infty$.

**Lemma 2.7.** (Puri and Ralescu [15]) The metric space $(\mathfrak{F}(\mathbb{R}), \hat{d})$ is complete.

For trapezoidal fuzzy numbers it is direct to verify that the following statements hold:

**Lemma 2.8.** a) If $\{(a_l^k, a_m^k, a_n^k, a_p^k) : 0 \le k \le M\}$ is a finite set of $M$ trapezoidal fuzzy numbers, then

$$\sum_{k=0}^{M}{}^* (a_l^k, a_m^k, a_n^k, a_p^k) = \left(\sum_{k=0}^{M} a_l^k, \sum_{k=0}^{M} a_m^k, \sum_{k=0}^{M} a_n^k, \sum_{k=0}^{M} a_p^k\right).$$

b) If $\{\widetilde{y}_k = (a_l^k, a_m^k, a_n^k, a_p^k) : k \geq 0\}$ is a sequence of trapezoidal numbers and $\sum_{k=0}^{\infty} a_i^k$ converges, $i \in \{l, m, n, p\}$, then

$$\left\{ \sum_{k=0}^{t}{}^{*} \widetilde{y}_k \right\}$$

converges when $t \longrightarrow \infty$ to the trapezoidal fuzzy number (see Definition 2.6):

$$\sum_{k=0}^{\infty}{}^{*} \widetilde{y}_k = \left( \sum_{k=0}^{\infty} a_l^k, \sum_{k=0}^{\infty} a_m^k, \sum_{k=0}^{\infty} a_n^k, \sum_{k=0}^{\infty} a_p^k \right).$$

Now, for $\widetilde{\eta}, \widetilde{\mu} \in \mathfrak{F}(\mathbb{R})$, with $\alpha$-cuts $\widetilde{\eta}_\alpha = [a_\alpha, b_\alpha]$ and $\widetilde{\mu}_\alpha = [c_\alpha, d_\alpha]$, $\alpha \in [0, 1]$, respectively, define $\widetilde{\eta} \leqslant^* \widetilde{\mu}$ if and only if $a_\alpha \leq c_\alpha$ and $b_\alpha \leq d_\alpha$ for all $\alpha \in [0, 1]$ (see [11]). It is not difficult to verify that the order "$\leq^*$" is, in fact, a partial order on $\mathfrak{F}(\mathbb{R})$.

**Remark 2.9.** Take $z_1, z_2 \in \mathbb{R}$, and let $\widetilde{z_1}$ and $\widetilde{z_2}$ be fuzzy numbers with membership functions given by $\widetilde{z_k}'(x)=1$, $x = z_k$ and $\widetilde{z_k}'(x) = 0$, $x \neq z_k$, $k = 1, 2$, respectively. Then, it is easy to see that $\widetilde{z_1} \leqslant^* \widetilde{z_2}$ is equivalent to $z_1 \leq z_2$.

Following [13] and [15], the next definitions on fuzzy random variables and their expectations are established. For this, $\mathfrak{C}(\mathbb{R})$ denotes the class of nonempty compact subsets of $\mathbb{R}$, and if $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces, then $\mathcal{A}_1 \otimes \mathcal{A}_2$ denotes the corresponding product $\sigma$-algebra associated to the product space $\Omega_1 \times \Omega_2$.

**Definition 2.10.** Let $(\Omega, \mathcal{A})$ be a measurable space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be the measurable space of the set of real numbers. A function $\widetilde{Y} : \Omega \longrightarrow \mathfrak{F}(\mathbb{R})$ is said to be a *fuzzy random variable* associated with $(\Omega, \mathcal{F})$, if the section $\widetilde{Y}_\alpha : \Omega \longrightarrow \mathfrak{C}(\mathbb{R})$ which is the $\alpha$-level function defined by $\widetilde{Y}_\alpha(\omega) = (\widetilde{Y}(\omega))_\alpha$ for all $\omega \in \Omega$ and $\alpha \in [0, 1]$ satisfies that $Gr(\widetilde{Y}_\alpha) = \{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in (\widetilde{Y}(\omega))_\alpha\} \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R})$, for all $\alpha \in [0, 1]$.

**Definition 2.11.** Given a probability space $(\Omega, \mathcal{A}, P)$, a fuzzy random variable $\widetilde{Y}$ associated to $(\Omega, \mathcal{A})$ is said to be an *integrably bounded fuzzy random variable* with respect to $(\Omega, \mathcal{A}, P)$ if there is a function $h : \Omega \longrightarrow \mathbb{R}$, $h \in L^1(\Omega, \mathcal{A}, P)$ such that $|x| \leq h(\omega)$, for all $(\omega, x) \in \Omega \times \mathbb{R}$ with $x \in (\widetilde{Y}(\omega))_0 := \widetilde{Y}_0(\omega)$.

**Definition 2.12.** Given an integrably bounded fuzzy random variable $\widetilde{Y}$ associated with respect to the probability space $(\Omega, \mathcal{A}, P)$, then the *fuzzy expected value* of $\widetilde{Y}$ in Aumann'sense is the unique fuzzy set of $\mathbb{R}$, $E^*[\widetilde{Y}]$ such that for each $\alpha \in [0, 1]$:

$$\left(E^*[\widetilde{Y}]\right)_\alpha = \left\{ \int_\Omega f(\omega)\, \mathrm{d}P(\omega) \mid f : \Omega \longrightarrow \mathbb{R}, f \in L^1(P), f(\omega) \in (\widetilde{Y}(\omega))_\alpha \ \ a.s.\ [P] \right\}. \quad (3)$$

## 3. MARKOV DECISION PROCESSES WITH TOTAL REWARD

Let $(X, A, \{A(i) \mid i \in X\}, P, R)$ be the usual discrete-time Markov decision model (see [16] and [19]), where both the *state space* $X$ and the *decision/control space* $A$ are finite sets. For each $i \in X$, $A(i) \subset A$, $A(i) \neq \varnothing$, is the subset of *admissible actions* in the state $i$. Let $\mathbb{K} := \{(i, a) \mid i \in X,\ a \in A(i)\}$. $P = [p_{ij}(a)]$ is the *controlled transition law* on $X$ given $\mathbb{K}$, and for each $(i, a) \in \mathbb{K}$, $p_{ij}(a) \geq 0$ and $\sum_{j \in X} p_{ij}(a) = 1$. Finally, the *reward per-stage* $R$ is a real-valued function on $\mathbb{K}$.

***Strategies***

A *decision strategy* $\pi$ is a (possibly randomized) rule for choosing actions, and at each time $t\,(t = 0, 1, \dots,)$ the decision prescribed by $\pi$ may depend on the current state as well as on the history of the previous states and actions. The set of all strategies will be denoted by $\Pi$. Given an initial state $i \in X$ and $\pi \in \Pi$ there is a canonical space $(\Omega, \mathcal{A}, P_{i,\pi})$ with the corresponding state-action process $\{(x_t, a_t)\}$ (for details, see [16]). $E_{i,\pi}$ denotes the expectation operator with respect to the probability measure $P_{i,\pi}$, and the stochastic process $\{x_t\}$ will be called *Markov decision process*. $\mathbb{F}$ denotes the set of functions $f : X \to A$ such that $f(i) \in A(i)$ for all $i \in X$. A strategy $\pi \in \Pi$ is *stationary* if there exists $f \in \mathbb{F}$ such that, under $\pi$, the action $f(x_t)$ is applied at each time $t$. The class of stationary strategies is naturally identified with $\mathbb{F}$.

***Optimality Criterion***

Given $\pi \in \Pi$ and initial state $x_0 = i \in X$, let

$$V(i, \pi) = E_{i,\pi}\Big[\sum_{t=0}^{\infty} R(x_t, a_t)\Big] \tag{4}$$

be the *expected total reward* when using the strategy $\pi$, given the initial state $i$. Furthermore, the $T$-stage reward is defined as follows:

$$V_T(i, \pi) := E_{i,\pi}\left[\sum_{t=0}^{T-1} R(x_t, a_t)\right]. \tag{5}$$

A strategy $\pi_o$ is said to be *optimal* if $V(i, \pi_o) = V_o(i)$ for all $i \in X$, where

$$V_o(i) = \sup_{\pi \in \Pi} V(i, \pi), \tag{6}$$

$i \in X$. $V_o$ defined in (6) is called the *optimal value function*.

**Assumption 3.1.** 1. For each $(i, a) \in \mathbb{K}$, $R(i, a) \geq 0$.

2. $V_o(i) < \infty$, for each $i \in X$.

**Lemma 3.2.** (Puterman [16], Ross [18])

a) The optimal value function $V_o$ satisfies the following *optimality equation*: for each $i \in X$,

$$V_o(i) = \sup_{a \in A(i)} \left[ R(i,a) + \sum_j p_{ij}(a) V_o(j) \right].$$

b) If $W : X \longrightarrow [0, \infty)$ satisfies that $W(i) \geq \sup_{a \in A(i)} \left[ R(i,a) + \sum_j p_{ij}(a) W(j) \right]$ for every $i \in X$, then $W \geq V_o$.

**Lemma 3.3.** ([5]) Under Assumption 3.1, there exists an optimal stationary policy $f_o$.

## 4. MARKOV DECISION PROCESSES WITH FUZZY TOTAL REWARD

Now, the new Markov decision model:

$$\left( X, A, \{A(i) \mid i \in X\}, P, \widetilde{R} \right), \tag{7}$$

will be analyzed. Note that this new Markov decision model has the same state and action spaces, the same restriction sets, and the same transition probability law as the Markov decision model described in Section 3. Hence, the sets of stationary and randomized policies coincide for both models, moreover, for each $i \in X$ and $\pi \in \Pi$ there is a canonical space $(\Omega, \mathcal{A}, P_{i,\pi})$ with the corresponding sequences $\{x_t\}$ and $\{a_t\}$ of states and decisions, respectively. Next, the objective function will be established, but firstly the fuzzy reward function $\widetilde{R}$ will be presented in the following assumption.

**Assumption 4.1.** Let $B$, $C$, $D$, and $F$ be nonnegative numbers such that: $0 \leq B < C \leq D < F$. It will be supposed that

$$\widetilde{R}(i,a) = R(i,a)\,(B, C, D, F)\,,$$

for all $i \in X$ and $a \in A(i)$, where $R : \mathbb{K} \longrightarrow \mathbb{R}$ is a reward function as it was considered in the previous section, and it is also supposed that Assumption 3.1 holds for the model $(X, A, \{A(i) \mid i \in X\}, P, R)$.

**Remark 4.2.** Observe that $\widetilde{0} \leqslant^* \widetilde{R}(i,a)$, for all $i \in X$ and $a \in A(i)$.

**Lemma 4.3.** Let $i \in X$ and $\pi \in \Pi$, and let $(\Omega, \mathcal{A}, P_{i,\pi})$ be the corresponding canonical space fixed (see Section 3). Let Y be a nonnegative discrete random variable associated to $(\Omega, \mathcal{A}, P_{i,\pi})$ such that $E_{i,\pi}[Y]$ exists. Then, $\widetilde{Y} = Y(B, C, D, F)$ is a fuzzy random variable associated to $(\Omega, \mathcal{A}, P_{i,\pi})$, and

$$E^*_{i,\pi}[\widetilde{Y}] = E_{i,\pi}[Y](B, C, D, F). \tag{8}$$

P r o o f . Take $i \in X$ and $\pi \in \Pi$. Let $Y$ be a nonnegative discrete random variable with finite or denumerable range denoted by $Y[\Omega] = \{y_1, y_2, \ldots\}$ and let $[Y = y_j] := \{\omega \in \Omega \mid Y(\omega) = y_j\}$, $j = 1, 2, \ldots$. Put $\Theta = (B, C, D, F)$ with $\alpha$-cuts $\Theta_\alpha = [q(\alpha), s(\alpha)]$, $\alpha \in [0, 1]$. Fix $\alpha \in [0, 1]$. Consider the following multifunction given by

$$\widetilde{Y}_\alpha(\omega) := (\widetilde{Y}(\omega))_\alpha = (Y(\omega)\Theta)_\alpha = Y(\omega)[q(\alpha), s(\alpha)],$$

$\omega \in \Omega$.
Now, notice that

$$\begin{aligned} Gr(\widetilde{Y}_\alpha) &= \{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in \widetilde{Y}_\alpha(\omega)\} \\ &= \{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in Y(\omega)[q(\alpha), s(\alpha)]\} \\ &= \bigcup_j ([Y = y_j] \times y_j[q(\alpha), s(\alpha)]). \end{aligned}$$

Hence, $Gr(\widetilde{Y}_\alpha) \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R})$. Since $\alpha$ is arbitrary, from Definition 2.10 it yields that $\widetilde{Y}$ is a fuzzy random variable. Next, note that, for each $\omega \in \Omega$,

$$(\widetilde{Y}(\omega))_0 := \widetilde{Y}_0(\omega) = Y(\omega)[B, F].$$

Define $h : \Omega \longrightarrow \mathbb{R}$ given by

$$h(\omega) := Y(\omega)F,$$

$\omega \in \Omega$. Then, trivially:

$$|x| \leq h(\omega),$$

$(\omega, x) \in \Omega \times \mathbb{R}$ with $x \in Y(\omega)[B, F]$. Also, clearly $E_{i,\pi}[h] = FE_{i,\pi}[Y]$ is finite. Therefore, from Definition 2.11, $\widetilde{Y}$ is an integrably bounded fuzzy random variable with respect to $(\Omega, \mathcal{A}, P_{i,\pi})$.

Now, from Definition 2.12, there is a unique fuzzy expected value $E^*_{i,\pi}[\widetilde{Y}]$ and it is direct to verify from (3) that, for each $\alpha$,

$$(E^*_{i,\pi}[\widetilde{Y}])_\alpha = E_{i,\pi}[Y][q(\alpha), s(\alpha)]$$

which, for each $\alpha$, is the $\alpha$-cut of the trapezoidal number given for

$$E_{i,\pi}[Y](B, C, D, F),$$

that is,

$$E^*_{i,\pi}[\widetilde{Y}] = E_{i,\pi}[Y](B, C, D, F).$$

□

**Lemma 4.4.** Suppose that Assumption 4.1 holds. Take $i \in X$ and $\pi \in \Pi$, and let $(\Omega, \mathcal{A}, P_{i,\pi})$ be the corresponding canonical space fixed. Then,

a) For each $T \geq 0$,

$$\widetilde{S}_T := \sum_{t=0}^{T} R\left(x_t, a_t\right)(B, C, D, F),$$

is a fuzzy random variable and

$$E^*_{i,\pi}\left[\widetilde{S}_T\right] = E_{i,\pi}\left[\sum_{t=0}^{T} R\left(x_t, a_t\right)\right](B, C, D, F).$$

b) Define:

$$H_{finite} = \left\{\omega \in \Omega \mid \sum_{t=0}^{\infty} R\left(x_t, a_t\right)(\omega) < +\infty\right\}$$

and

$$H_{\infty} = \left\{\omega \in \Omega \mid \sum_{t=0}^{\infty} R\left(x_t, a_t\right)(\omega) = +\infty\right\}.$$

Then, $\widetilde{S}$ given by

$$\widetilde{S}(\omega) = \begin{cases} \sum_{t=0}^{\infty} R\left(x_t, a_t\right)(\omega)(B, C, D, F), & \omega \in H_{finite} \\ \widetilde{0}, & \omega \in H_{\infty} \end{cases} \tag{9}$$

is a fuzzy random variable, and

$$E^*_{i,\pi}[\widetilde{S}] = E_{i,\pi}\left[\sum_{t=0}^{\infty} R\left(x_t, a_t\right)\right](B, C, D, F). \tag{10}$$

Proof.

a) Notice that for each $T \geq 0$, $\sum_{t=0}^{T} R\left(x_t, a_t\right)$ is a nonnegative discrete random variable (recall that $X$ and $A$ are finite sets). Consequently, this part a) follows from Lemma 4.3, with $Y = \sum_{t=0}^{T} R\left(x_t, a_t\right)$.

b) Denote, for each $T \geq 0$, $S_T := \sum_{t=0}^{T} R\left(x_t, a_t\right)$, where its finite range is given by

$$S_T[\Omega] = \{y_1^T, y_2^T, \ldots, y_{k_T}^T\},$$

and consider the measurable sets $[S_T = y_j^T] := \{\omega \in \Omega \mid Y(\omega) = y_j^T\}$, $j = 1, 2, \ldots, k^T$. Let $S = \sum_{t=0}^{\infty} R\left(x_t, a_t\right)$.

Notice that from Assumption 4.1, $0 \leq E_{i,\pi}[S] < \infty$ which implies that $S$ is finite a.s. $[P_{i,\pi}]$ (see Exercise 4Q, p. 39 in [3]), that is, the measurable set $H_{\infty}$ satisfies that $P_{i,\pi}(H_{\infty}) = 0$. Now, from Lemma 2.8 in [3] it follows that

$$\mathring{S}(\omega) = \begin{cases} S(\omega), & \omega \in H_{finite} \\ 0, & \omega \in H_{\infty}, \end{cases}$$

is measurable, and with this (9) is established as $\widetilde{S}(\omega) = \mathring{S}(\omega)(B, C, D, F)$, $\omega \in \Omega$, that is

$$\widetilde{S}(\omega) = \begin{cases} S(\omega)(B, C, D, F), & \omega \in H_{finite} \\ \widetilde{0}, & \omega \in H_{\infty}. \end{cases}$$

Observe that, for $\omega \in H_{finite}$, i. e. if $S(\omega) < \infty$ it results that

$$\lim_{t \longrightarrow \infty} R(x_t(\omega), a_t(\omega)) = 0,$$

and recalling that $X$ and $A$ are finite sets and $R \geq 0$, it follows that there is a positive integer $\tau = \tau(\omega)$ such that

$$R(x_t(\omega), a_t(\omega)) = 0,$$

for all $t > \tau$, or

$$S(\omega) = S_{\tau}(\omega), \tag{11}$$

and in this case it also holds that:

$$\widetilde{S}(\omega) = S_{\tau}(\omega)(B, C, D, F).$$

Now, take into account the multifunction given by

$$\widetilde{S}_{\alpha}(\omega) := (\widetilde{S}(\omega))_{\alpha},$$

$\omega \in \Omega$.

Firstly, if $\omega \in \Omega \setminus H_{\infty}$ then, for each $\alpha$,

$$\widetilde{S}_{\alpha}(\omega) = (\mathring{S}(\omega)\Theta)_{\alpha} = S(\omega)[q(\alpha), s(\alpha)],$$

(recall that $\Theta = (B, C, D, F)$ with $\alpha$-cuts $\Theta_{\alpha} = [q(\alpha), s(\alpha)]$, $\alpha \in [0, 1]$). And, from (11),

$$\omega \in \bigcup_{j=1}^{k_{\tau}} [S_{\tau} = y_j^{\tau}].$$

Thus, if $x \in \widetilde{S}_{\alpha}(\omega) = S(\omega)[q(\alpha), s(\alpha)]$ then

$$(\omega, x) \in \left[ \bigcup_{T=0}^{+\infty} \bigcup_{j=1}^{k_T} ([S_T = y_j^T] \times y_j^T [q(\alpha), s(\alpha)]) \right].$$

Secondly, if $\omega \in H_{\infty}$, then for each $\alpha$,

$$\widetilde{S}_{\alpha}(\omega) = (\tilde{0})_{\alpha} = \{0\},$$

hence, if $x \in \{0\}$ it results that

$$(\omega, x) \in H_{\infty} \times \{0\}.$$

In conclusion,

$$Gr(\widetilde{S}_\alpha) = \left\{(\omega, x) \in \Omega \times \mathbb{R} \mid x \in \widetilde{S}_\alpha(\omega)\right\}$$
$$\subseteq \left[\bigcup_{T=0}^{+\infty}\bigcup_{j=1}^{k_T}([S_T = y_j^T] \times y_j^T[q(\alpha), s(\alpha)])\right] \cup [H_\infty \times \{0\}].$$

On the other hand, it is direct to verify that

$$\left[\bigcup_{T=0}^{+\infty}\bigcup_{j=1}^{k_T}([S_T = y_j^T] \times y_j^T[q(\alpha), s(\alpha)])\right] \cup [H_\infty \times \{0\}] \subseteq Gr(\widetilde{S}_\alpha).$$

Consequently,

$$Gr(\widetilde{S}_\alpha) = \left[\bigcup_{T=0}^{+\infty}\bigcup_{j=1}^{k_T}([S_T = y_j^T] \times y_j^T[q(\alpha), s(\alpha)])\right] \cup [H_\infty \times \{0\}].$$

Therefore, $Gr(\widetilde{S}_\alpha) \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R})$. Since $\alpha$ is arbitrary, from Definition 2.10 it yields that $\widetilde{S}$ is a fuzzy random variable. And a proof similar to the one given in Lemma 4.3 to verify formula (8) allows to validate (10).

□

The following lemma is a direct consequence of Lemmas 2.8 and 4.4.

**Lemma 4.5.** Suppose that Assumption 4.1 holds. Take $i \in X$ and $\pi \in \Pi$, and let $(\Omega, \mathcal{A}, P_{i,\pi})$ be the corresponding canonical space fixed. Then, for each $T \geq 0$,

$$\widetilde{S}_T = \sum_{t=0}^{T}{}^{*} \widetilde{R}(x_t, a_t) = \sum_{t=0}^{T} R(x_t, a_t)(B, C, D, F),$$

and

$$E_{i,\pi}^{*}\left[\sum_{t=0}^{T}{}^{*} \widetilde{R}(x_t, a_t)\right] = E_{i,\pi}\left[\sum_{t=0}^{T} R(x_t, a_t)\right](B, C, D, F).$$

Moreover,

$$\widetilde{S} = \sum_{t=0}^{\infty}{}^{*} \widetilde{R}(x_t, a_t) = \begin{cases} \sum_{t=0}^{\infty} R(x_t, a_t)(B, C, D, F) \\ \tilde{0}, \end{cases} \tag{12}$$

and

$$E_{i,\pi}^{*}\left[\sum_{t=0}^{\infty}{}^{*} \widetilde{R}(x_t, a_t)\right] = E_{i,\pi}\left[\sum_{t=0}^{\infty} R(x_t, a_t)\right](B, C, D, F).$$

**Remark 4.6.** a) The (degenerate) case in which in the decision model (7), $\widetilde{R}(i,a)$ has a membership function given by:

$$(\widetilde{R}(i,a))'(x) = \begin{cases} 1 & \text{if } x = R(i,a) \\ 0 & \text{if } x \neq R(i,a), \end{cases}$$

for all $i \in X$ and $a \in A(i)$ implies that $\widetilde{R}$ is a fuzzy random variable and

$$E^*_{i,\pi}[\widetilde{R}(x_t,a_t)] = E_{i,\pi}[R(x_t,a_t)]\widetilde{1},$$

for all $i \in X$, $\pi \in \Pi$, and $t \geq 0$.

b) Note that Lemmas 4.4 and 4.5 validate all the fuzzy random variables and their expectations of this paper.

## 5. OPTIMAL CONTROL PROBLEM FOR THE FUZZY MODEL

**Definition 5.1.** For each $i \in X$ and $\pi \in \Pi$ the *fuzzy total reward* is defined by (12), and in this case the corresponding *fuzzy expectation* is given by:

$$\widetilde{V}(i,\pi) := E^*_{i,\pi}\left[\sum_{t=0}^{\infty}{}^{*}\, \widetilde{R}\,(x_t,a_t)\right] = E_{i,\pi}\left[\sum_{t=0}^{\infty} R\,(x_t,a_t)\right](B,C,D,F). \tag{13}$$

Now, let $i \in X$ and $\pi \in \Pi$, and $T \geq 0$:

$$\tilde{V}_T(i,\pi) := \sum_{t=0}^{T}{}^{*}\, E^*_{i,\pi}\left[\widetilde{R}\,(x_t,a_t)\right], \tag{14}$$

$V_T$ is known as the *$T$-stage fuzzy total reward.*

**Remark 5.2.** Observe that the *$T$-stage fuzzy total reward* (see (14)) is a trapezoidal fuzzy number, specifically,

$$\widetilde{V}_T(i,\pi) = (BV_T(i,\pi), CV_T(i,\pi), DV_T(i,\pi), FV_T(i,\pi))\,,$$

for $\pi \in \Pi$ and $i \in X$, where $V_T$ is the $T$-stage total crisp reward (see (5)).

**Lemma 5.3.** Suppose that Assumption 4.1 holds. Then, for each $i \in X$ and $\pi \in \Pi$, $\{\widetilde{V}_T(i,\pi)\}_{T=0}^{+\infty}$ converges and

$$\begin{aligned} \widetilde{V}(i,\pi) &:= lim^*_{T\to\infty}\widetilde{V}_T(i,\pi) \\ &= \sum_{t=0}^{\infty}{}^{*}\, E^*_{i,\pi}\left[\widetilde{R}\,(x_t,a_t)\right] \\ &= (BV(i,\pi), CV(i,\pi), DV(i,\pi), FV(i,\pi)), \end{aligned} \tag{15}$$

where

$$V(i,\pi) = \sum_{t=0}^{\infty} E_{i,\pi}\left[R\,(x_t,a_t)\right] \in \mathbb{R}.$$

P r o o f . Let $\pi \in \Pi$ and $i \in X$ be fixed. To simplify the notation, let us denote "$V(i,\pi)$" by "$V$" and "$V_T(i,\pi)$" by "$V_T$". Then, the $\alpha$-cut of (14), is given by

$$\begin{aligned}\Delta^T :&= (BV_T, CV_T, DV_T, FV_T)_\alpha \\ &= [B(1-\alpha)V_T + \alpha CV_T, F(1-\alpha)V_T + \alpha DV_T].\end{aligned}$$

Analogously,

$$\begin{aligned}\Delta :&= (BV, CV, DV, FV)_\alpha \\ &= [B(1-\alpha)V + \alpha CV, F(1-\alpha)V + \alpha DV].\end{aligned}$$

Hence, by (2), it is obtained that

$$\hat{d}(\Delta^T, \Delta) = \sup_{\alpha\in[0,1]} d(\Delta^T, \Delta).$$

Now, due to the identity $\max(c,b) = (c+b+|b-c|)/2$ with $b, c \in \mathbb{R}$, it yields that

$$d(\Delta^T, \Delta) = (1-\alpha)D(V - V_T) + \alpha C(V - V_T).$$

Then,

$$\begin{aligned}\hat{d}(\Delta^T, \Delta) &= \sup_{\alpha\in[0,1]} (V - V_T)(D - \alpha(D-C)) \\ &= (V - V_T)D.\end{aligned} \tag{16}$$

Therefore, when $T$ goes to infinity in (16), it concludes that

$$\begin{aligned}\lim_{T\longrightarrow\infty} \hat{d}(\widetilde{V}_T, \widetilde{V}) &= \lim_{T\longrightarrow\infty} (V - V_T)D \\ &= 0.\end{aligned}$$

The second equality is a consequence of (4) and (5). □

Now, the *fuzzy optimal control problem* is as follows: determine $\pi_o \in \Pi$ (if it exists) such that:

$$\widetilde{V}(i,\pi) \leq^* \widetilde{V}(i,\pi_o), \tag{17}$$

for all $i \in X$ and $\pi \in \Pi$. In this case it is possible to write

$$\widetilde{V}(i,\pi_o) = sup^*_{\pi\in\Pi} \widetilde{V}(i,\pi), \tag{18}$$

$i \in X$ and it is said that $\pi_o$ is *optimal*. Moreover, the function $\widetilde{V_o}(i) = \widetilde{V}(i,\pi_o)$ for $i \in X$ will be called the *fuzzy optimal value function*.

**Remark 5.4.** Using Remarks 2.9 and 4.6 a) it is direct to see that in the (degenerate) case when in the decision model (7), $\widetilde{R}(i,a)$ has a membership function given by:

$$(\widetilde{R}(i,a))'(x) = \begin{cases} 1 & \text{if } x = R(i,a) \\ 0 & \text{if } x \neq R(i,a), \end{cases}$$

for all $i \in X$ and $a \in A(i)$, then the fuzzy optimal control problem described in (17) and (18) is reduced to the optimal control problem described in (6).

**Lemma 5.5.** Suppose that Assumption 4.1 holds. Then, for each $i \in X$, $\widetilde{V}_o(i)$ is a bounded function, i. e. there exists $\widetilde{K} \in \mathcal{F}(\mathbb{R})$ such that $\widetilde{V}_o(i) \leq^* \widetilde{K}$, $i \in X$.

P r o o f . Take $i \in X$ fixed. Then, as a consequence of (15), the $\alpha$-cut of $\widetilde{V}(i,\pi)$ is given by

$$\widetilde{V}(i,\pi)_\alpha = [BV(i,\pi) + \alpha V(i,\pi)(C-B), FV(i,\pi) - \alpha V(i,\pi)(F-D)],$$

for each $\pi \in \Pi$. Note that because $X$ is finite, it is possible to take $K$ in order to $V(i,\pi) \leq K$, for all $i \in X$ and $\pi \in \Pi$. In consequence, observe that for each $\pi \in \Pi$ :

$$BV(i,\pi) + \alpha V(i,\pi)(C-B) \leq BK + \alpha(C-B)K,$$

and

$$\begin{aligned} FV(i,\pi) - \alpha V(i,\pi)(F-D) &\leq FK(1-\alpha) + \alpha DK \\ &= FK - \alpha(F-D)K. \end{aligned}$$

Consequently, $\widetilde{V}(i,\pi) \leqslant^* \widetilde{K} := (BK, CK, DK, FK)$, for each $\pi \in \Pi$. Therefore, $\widetilde{V}_o(i) \leqslant^* \widetilde{K}$ (see (18)). Since $i$ is arbitrary, the result follows.

□

**Theorem 5.6.** Under Assumption 4.1, the following statements hold.

a) The optimal policy of the fuzzy control problem is the same as the optimal policy of the standard optimal control problem.

b) The optimal fuzzy value function is given by

$$\widetilde{V}_o(i) = (BV_o(i), CV_o(i), DV_o(i), FV_o(i)), i \in X.$$

P r o o f .

a) Let $\pi \in \Pi$ and $i \in X$ be fixed. First, observe that (13) is equivalent to

$$\widetilde{V}(i,\pi) = (BV(i,\pi), CV(i,\pi), DV(i,\pi), FV(i,\pi)),$$

as a consequence of Assumption 4.1. Then, the $\alpha$-cut of $\tilde{V}(\pi,x)$ is given by

$$\widetilde{V}(\pi,x)_\alpha = [(C-B)V(i,\pi)\alpha + BV(i,\pi), FV(i,\pi) - (F-C)\alpha V(i,\pi)].$$

Now, by Lemma 3.3, there exists $f_o \in \mathbb{F}$ such that

$$(C-B)V(i,\pi)\alpha + BV(i,\pi) \leq (C-B)V(i,f_o)\alpha + BV(i,f_o).$$

and

$$FV(i,\pi)\alpha - (F-C)\alpha V(i,\pi) \leq FV(i,f_o)\alpha - (F-C)V(i,f_o).$$

and since $i \in X$ and $\pi \in \Pi$ are arbitrary, the result follows due to (18).

b) By the part a) of this theorem, it follows that

$$\widetilde{V}_o(i) = (BV(i,f_o), CV(i,f_o), DV(i,f_o), FV(i,f_o)),$$

for each $i \in X$; hence, it is concluded that

$$\widetilde{V}_o(i) = (BV_o(i), CV_o(i), DV_o(i), FV_o(i)), i \in X.$$

□

## 6. EXAMPLES

### 6.1. An optimal stopping problem

Section 7.2.8 in [16] presents several optimal stopping problems seen as total reward MDPs. Here, a similar version of Example 7.2.6 in [16] and its extension to fuzzy environment is provided.

Consider a finite Markov chain with state space $X^{'} = \{i_1, i_2, i_3, i_4\}$, and with transition probability matrix:

$$P = \begin{bmatrix} 0 & 1/3 & 2/3 & 0 \\ 4/5 & 1/5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

That is, each entry of $P$ describes the transition probability $p_{ij}$, for $i, j \in X^{'}$. Suppose that at each decision epoch, the controller has two admissible actions: to stop(Q) or to continue(C). For each $i \in X^{'}$, $A(i) = \{C, Q\}$. Now, choosing $C$ in state $i$ causes the system to move to state $j \in X^{'}$ with probability $p_{ij}$, and choosing $Q$ moves the system to state $\delta$, at which it receives no subsequent rewards. Notice that $X = \{i_1, i_2, i_3, i_4, \delta\}$ and $A(\delta) = \{C\}$. In particular, suppose that $R(i_1, Q) = g(i_1) = 8$, $R(i_2, Q) = g(i_2) = 5$, $R(i_3, Q) = g(i_3) = 3$, $R(i_4, Q) = g(i_4) = 0$, and $R(\delta, C) = 0$.

The objective is to determine a policy maximizing the expected total reward, under the assumption that the rewards are received only upon termination. Then, due to Theorem 7.2.3 (a) in [16], the optimal value function $V_o$ is the minimal solution in the class of functions $w : X^{'} \longrightarrow \mathbb{R}$ with $w \geq 0$ of the following dynamic programming equation:

$$w(i) = \max\{g(i), \sum_{j \in X'} w(j) p_{ij}\}. \tag{19}$$

$i \in X^{'}$ with $V_o(\delta) = 0$. Then, applying the linear programming approach, (19) is equivalent to the following linear program:

$$\text{MINIMIZE} : w(i_1) + w(i_2) + w(i_3) + w(i_4)$$

subject to

$$w(i) \geq g(i), \tag{20}$$

$$w(i) \geq \sum_{j \in X'} w(j) p_{ij}, \tag{21}$$

$i \in X^{'}$. Then, inequalities (20) and (21) are equivalent to

$$\begin{aligned} 3w(i_1) - w(i_2) - 2w(i_3) &\geq 0, \\ w(i_2) - w(i_1) &\geq 0, \\ 2w(i_3) - w(i_1) - w(i_4) &\geq 0, \end{aligned}$$

$$w(i_1) \geq 8, \;\; w(i_2) \geq 5, \;\; w(i_3) \geq 3, \;\; w(i_4) \geq 0.$$

Applying the simplex algorithm, it is obtained that the optimal solution is $V_o(i_1) = 8$, $V_o(i_2) = 8$, $V_o(i_3) = 4$ and $V_o(i_4) = 0$. Consequently, from Theorem 7.2.22 b) in [16], the optimal stationary policy $f_o$ is given by:

$$f_o(i) = \begin{cases} Q & \text{if} \quad i \in \{i_1, i_4\} \\ C & \text{if} \quad i \in \{i_2, i_3\}. \end{cases} \tag{22}$$

Now, consider the following trapezoidal fuzzy reward function:

$$\widetilde{R}(i, Q) = (0, \frac{9}{10}R(i, Q), \frac{11}{10}R(i, Q), 2R(i, Q)),$$

$i \in X^{'}$ with the interpretation that the trapezoidal number $(0, 0, 0, 0)$ is equal to $\widetilde{0}$, and $\widetilde{R}(\delta, C) = \widetilde{0}$.

Concretely, for the decision $Q$ the fuzzy rewards are given by:

- $\widetilde{R}(i_1, Q) = (0, 7.2, 8.8, 16)$,
- $\widetilde{R}(i_2, Q) = (0, 4.5, 5.5, 10)$,
- $\widetilde{R}(i_3, Q) = (0, 2.7, 3.3, 6)$,
- $\widetilde{R}(i_4, Q) = \widetilde{0}$.

**Remark 6.1.** Note that, for instance, $\widetilde{R}(i_1, Q) = (0, 7.2, 8.8, 16)$ models the fact that in state $i_1$, the reward received only upon termination is approximately in the interval $[7.2, 8.8]$ instead of receiving the exact quantity of $g(i_1) = 8$ in the standard MDP; the rest of the fuzzy rewards have a similar interpretation.

Now, applying Theorem 5.6, the optimal policy of the fuzzy control problem is the same as the optimal policy $f_o$ of the optimal control problem given in (22), and the optimal fuzzy value function is given by:

$$\widetilde{V}_o(i) = (0, \frac{9}{10}V_o(i), \frac{11}{10}V_o(i), 2V_o(i)),$$

$i \in X^{'}$. And,

$$\widetilde{V}_o(\delta) = 0.$$

Consequently,

- $\widetilde{V}_o(i_1) = (0, 7.2, 8.8, 16)$,
- $\widetilde{V}_o(i_2) = (0, 7.2, 8.8, 16)$,
- $\widetilde{V}_o(i_3) = (0, 3.6, 4.4, 8)$,
- $\widetilde{V}_o(i_4) = \widetilde{0}$.

### 6.2. A red-black model

The first part of this subsection is based on [19], pp. 76–83, and later the fuzzy extension is provided.

An individual possessing $i$ dollars enters a gambling casino that allows any bet as follows: If you possess $i$ dollars, then you are allowed to bet any positive integer less than or equal to $i$. Furthermore, if you bet $j$ then you either

(a) win $j$ with probability $p$ or

(b) lose $j$ with probability $1-p$.

The question established in [19] is: What gambling strategy maximizes the probability that the individual will attain a fortune of $N$ before going broke? The answer to this question fits the framework of the MDPs with the total reward given in Section 3, where the state is the gamblers' fortunes, since if it is supposed that a terminal reward of 1 is earned if we ever reach the state $N$ and all other rewards are zero, then the expected total reward equals the probability of ever reaching state $N$. Specifically, this gambling model is formulated as follows:

*Description of the Model*

(a) $X = \{0, 1, \ldots, N\}$, where we say that the state is $i$ when the present fortune is $i$.

(b) Let $[k]$ be the integer part of $k$. If the present fortune is $i$, then it would never pay to bet more than $N - i$, that is,

$$A = \{0, 1, \ldots, [N/2]\},\ A(0) = \{0\},\ A(i) = \{1, 2, \ldots, \min\{i, N-i\}\},\ i \neq 0.$$

(c) $p_{i,i+a}(a) = p$, $p_{i,i-a}(a) = 1-p$, $p_{N0}(a) = 1$, $p_{00}(0) = 1$.

(d) $R(i,a) = 0$, $i \neq N$, $a \in A(i)$, and $R(N,0) = 1$.

**Remark 6.2.** Let $\Xi$ be the set of ever reaching the state $N$. Notice that, for each strategy $\pi \in \Pi$ and $i \in X$, $V(i,\pi) = P_{i,\pi}(\Xi)$ (see [19]).

Define the *timid strategy* $\tau$ to be that strategy which always bets 1, and define the *bold strategy* $\beta$ to be the strategy that, if the present fortune is $i$,

(a) bets $i$ if $i \leq \dfrac{N}{2}$,

(b) bets $N - i$ if $i \geq \dfrac{N}{2}$

From Proposition 2.1 and Corollary 2.6 in [19] the following lemma is obtained.

**Lemma 6.3.**

(a) If $p \geq \frac{1}{2}$, then $\tau$ maximizes the probability of ever attaining a fortune $N$, i. e., in this case, $V_o(i) = V(i, \tau)$, for all $i \in X$.

(b) If $p \leq \frac{1}{2}$, then $\beta$ maximizes the probability of ever attaining a fortune $N$, i. e., in this case, $V_o(i) = V(i, \beta)$, for all $i \in X$.

Now, applying Theorem 5.6, the result on the fuzzy red-black model will be presented.

**Theorem 6.4.** Suppose that Assumption 4.1 holds.

(a) If $p \geq \frac{1}{2}$, then $\widetilde{V}(i, \pi) \leq^* \widetilde{V}(i, \tau)$, for all $\pi \in \Pi$ and $i \in X$. Therefore, $\tau$ is optimal and

$$\widetilde{V}(i, \tau) = (BV(i, \tau), CV(i, \tau), DV(i, \tau), FV(i, \tau)),$$

$i \in X$.

(b) If $p \leq \frac{1}{2}$, then $\widetilde{V}(i, \pi) \leq^* \widetilde{V}(i, \beta)$, for all $\pi \in \Pi$ and $i \in X$. Therefore, $\beta$ is optimal and

$$\widetilde{V}(i, \beta) = (BV(i, \beta), CV(i, \beta), DV(i, \beta), FV(i, \beta)),$$

$i \in X$.

**Remark 6.5.** Observe that in the non-fuzzy red-black model the gambler's goal is to reach at the end of the game a certain fortune $N$. Now, following the description of the nonfuzzy red-black model and Assumption 4.1 it is obtained that for the fuzzy model: $\widetilde{R}(i, a) = \widetilde{0}$, $i \neq N$, $a \in A(i)$, and $\widetilde{R}(N, 0) = (B, C, D, F)$; therefore, taking $C \leq N \leq D$, it could be interpreted that the gambler receives at the end of the game a quantity between the bounds $C$ and $D$ instead of the gambler getting the exact quantity $N$ as in the nonfuzzy model.

## 7. CONCLUSIONS

The theory presented in this article takes into account imprecision or ambiguity in the reward function resulting in an MDPs theory on finite spaces with a fuzzy total reward function. This extends the standard MDPs theory on finite spaces and the total reward criterion. And, with the assumptions for the fuzzy optimal problem provided in the article, the main consequences obtained are that the optimal policy of the fuzzy problem coincides with the standard non-fuzzy problem and the fuzzy optimal value function is of a convenient trapezoidal form. It is relevant to remark that, in the fuzzy version of the gambling model given, the bold and the timid strategies, which are well-known in the gambling context, appear as the optimal strategies for the player, and that the fortune $N$ which at the end of the game the player will receive can be substituted by the fact that $N$ belongs "approximately in a certain interval".

## ACKNOWLEDGEMENT

R. Montes-de-Oca thanks Prof. Miguel López-Díaz for fruitful comments on fuzzy random variables, and kindly providing some references on this topic, particularly reference [13] of this article.

(Received December 2, 2021)

## REFERENCES

[1] S. Abbasbandy and T. Hajjari: A new approach for ranking of trapezoidal fuzzy numbers. Comput. Math. Appl. *57* (2009), 413–419. DOI:10.1016/j.camwa.2008.10.090

[2] A. I. Ban: Triangular and parametric approximations of fuzzy numbers inadvertences and corrections. Fuzzy Sets and Systems *160* (2009), 3048–3058. DOI:10.1016/j.fss.2009.04.003

[3] R. G. Bartle: The Elements of Integration. Wiley, New York 1995.

[4] R. E. Bellman and L. A. Zadeh: Decision-making in a fuzzy enviroment. Management Sci. *17* (1970), 141–164. DOI:10.1287/mnsc.17.4.B141

[5] R. Cavazos-Cadena and R. Montes-de-Oca: Existence of optimal stationary policies in finite dynamic programs with nonnegative rewards. Probab. Engrg. Inform. Sci. *15* (2001), 557–564. DOI:10.1017/s0269964801154082

[6] S. H. Chen: Operations of fuzzy numbers with step form membership function using function principle. Information Sci. *108* (1998), 149–155. DOI:10.1016/S0020-0255(97)10070-6

[7] P. Diamond and P. Kloeden: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific, Singapore 1994.

[8] D. Driankov, H. Hellendoorn, and M. Reinfrank: An Introduction to Fuzzy Control. Springer Science and Business Media, New York 2013.

[9] R. Efendi, N. Arbaiy, and M. M. Deris: A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. Information Sci. *441* (2018), 113–132. DOI:10.1016/j.ins.2018.02.016

[10] M. Fakoor, A. Kosari, and M. Jafarzadeh: Humanoid robot path planning with fuzzy Markov decision processes. J. Appl. Res. Tech. *14* (2016), 300–310. DOI:10.1016/j.jart.2016.06.006

[11] N. Furukawa: Parametric orders on fuzzy numbers and their roles in fuzzy optimization problems. Optimization *40* (1997), 171–192. DOI:10.1080/02331939708844307

[12] M. Kurano, M. Yasuda, J. Nakagami, and Y. Yoshida: Markov decision processes with fuzzy rewards. In: Proc. Int. Conf. on Nonlinear Analysis, Hirosaki 2002, pp. 221–232.

[13] M. López-Díaz and D. A. Ralescu: Tools for fuzzy random variables: embeddings and measurabilities. Comput. Statist. Data Anal. *51* (2006), 109–114. DOI:10.1016/j.csda.2006.04.017

[14] W. Pedrycz: Why triangular membership functions?. Fuzzy Sets and Systems *64* (1994), 21–30. DOI:10.1016/0165-0114(94)90003-5

[15] M. L. Puri and D. A. Ralescu: Fuzzy random variable. J. Math. Anal. Appl. *114* (1986), 402–422. DOI:10.1016/0022-247x(86)90093-4

[16] M. L. Puterman: Markov Decision Processes: Discrete Stochastic Dynamic. First edition. Wiley–Interscience, California 2005.

[17] S. Rezvani and M. Molani: Representation of trapezoidal fuzzy numbers with shape function. Ann. Fuzzy Math. Inform. *8* (2014), 89–112.

[18] S. Ross: Dynamic programming and gambling models. Adv. Appl. Probab. *6* (1974), 593–606. DOI:10.1017/S0001867800040027

[19] S. Ross: Introduction to Stochastic Dynamic Programming. Academic Press, New York 1983.

[20] A. Semmouri, M. Jourhmane, and Z. Belhallaj: Discounted Markov decision processes with fuzzy costs. Ann. Oper. Res. *295* (2020), 769–786. DOI:10.1007/s10479-020-03783-6

[21] A. Syropoulos and T. Grammenos: A Modern Introduction to Fuzzy Mathematics. Wiley, New Jersey 2020.

[22] L. Zadeh: Fuzzy sets. Inform. Control *8* (1965), 338–353. DOI:10.1016/S0019-9958(65)90241-X

[23] W. Zeng and H. Li: Weighted triangular approximation of fuzzy numbers. Int. J. Approx. Reason. *46* (2007), 137–150. DOI:10.1017/S0012217300001591

*Karla Carrero-Vera, Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y Río Verde, Col. San Manuel, CU, Puebla, Pue. 72570. México.*

*e-mail: karla.carrero@alumno.buap.mx*

*Hugo Cruz-Suárez, Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y Río Verde, Col. San Manuel, CU, Puebla, Pue. 72570. México.*

*e-mail: hcs@fcfm.buap.mx*

*Raúl Montes-de-Oca, Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina 09340. México City. México.*

*e-mail: momr@xanum.uam.mx*