

A fixed-shape fuzzy median of a fuzzy sample

Adrian Ban,¹ Lucian Coroianu,¹ Przemysław Grzegorzewski²

¹Department of Mathematics and Informatics, University of Oradea,
Universităţii 1, 410087 Oradea, Romania

²Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland

Abstract

Since fuzzy numbers are not linearly ordered a median of a fuzzy sample cannot be determined directly as it happens with crisp observations. Even the very notion of a median of a fuzzy sample is not well established. Therefore we propose some definitions of the fixed-shape fuzzy median of a fuzzy sample as a member of a given class of fuzzy numbers which minimizes the distance to the sample. We also prove the existence of such fixed-shape median for the most important subfamilies of fuzzy numbers and the broad class of metrics.

Keywords: Fuzzy numbers, median, distances between fuzzy numbers

1. Introduction

Different summary measures are applied in descriptive statistics to characterize briefly a data set under study. One can distinguish three groups of descriptive measures: measures of location, measures of dispersion and measures of shape. A subfamily of location measures, called *central tendency measures*, seems to be the most important one since it characterizes values typical for a sample. There are many well-known central tendency measures but the most popular and the most important one is the *average*, also called the *arithmetic mean*. It is so because the average, considered as an estimator of the population mean, possesses many desired properties: it is unbiased and consistent. Moreover, it is efficient for many typical distributions.

But even though the average cannot be perceived as the overall champion among central tendency measures in any possible real-life situations since it has a strong drawback - it is quite sensitive to outliers. Indeed, observations that are too big or too small with respect to the majority of observations in a sample may have a serious impact on the final value of the average. Due to such undesired behavior the average is referred as a statistical tool nonrobust to outliers.

To avoid this unpleasant effects of outliers one may apply a *statistical median* as a central tendency measure. Since median is defined as a middle element in the ordered sequence of observations (if a sample size is odd) or as the average of the two middle observations (if a sample size is even), it is -

by definition - not influenced by the extreme values and so it is perfectly robust to outliers. Of course, since the median is based only on the middle of a sample it does not utilize so much information on the underlying sample as the average and therefore it cannot be in general regarded as a better location characteristic than the average unless the presence of outliers.

All these considerations and problems well-known from the classical descriptive statistics holds for fuzzy data analysis. Actually, having a fuzzy sample that describes imprecise observations, we are also interested in finding summary measures that characterize the most important properties of the data set in a concise way. Hence we need reasonable tools to aggregate fuzzy sets, like fuzzy average to characterize a typical value for a fuzzy sample. And here we immediately notice that a fuzzy counterpart of the average is also sensitive to outliers, no matter crisp or fuzzy, as its classical prototype. If so, it is not surprising that a natural idea of a statistical tool which may be robust to outliers in a fuzzy environment would be a *fuzzy median*. Unfortunately, when we move from the real to fuzzy data and want to generalize the concept of the median to fuzzy context, we see immediately that it is not possible since fuzzy numbers do not form a linear order and so we cannot find the “middle” element in a sample. Does it mean that we have to abandon the idea of the median in a fuzzy environment? Fortunately not, but a fuzzy generalization of the median may be defined in several ways (see, e.g. [3, 5, 11, 12, 13]) and its existence and properties are neither obvious nor evident.

The paper is organized as follows: in Sec. 2 we recall basic notions connected with the crisp median known from the classical data analysis and with fuzzy numbers applied for modelling imprecise observations. Moreover, an idea of a fuzzy median proposed in [5] is also given. Next, in Sec. 3, we discuss problems connected with the determination of the median of a fuzzy sample and we suggest the idea of the fixed-shape fuzzy median. In Sec. 4 we mention basic metrics applied in a family of fuzzy numbers. In Sec. 5 we cite and show some important mathematical facts necessary for our further considerations, while in Sec. 6 we present some existence results in approximation problems. Finally, in Sec. 7 we prove the existence of the fuzzy median

of a fuzzy sample under very general assumptions on a fuzzy environment.

Because of the lack of space we had to omit most of the proofs of the theorems and corollaries needed for the main result.

2. Basic concepts and notation

Let X_1, \dots, X_n denote a usual crisp sample made up by real numbers. Then the sample median $Med = Med(X_1, \dots, X_n)$ is defined as a middle observation - if the sample size is odd, or as the average from the two middle observations - if n is even, i.e.

$$Med = \begin{cases} X_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{1}{2} (X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}) & \text{if } n \text{ is even,} \end{cases}$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ denote order statistics from the sample X_1, \dots, X_n .

Now let A_1, \dots, A_n denote a fuzzy sample of imprecise observations, where each A_i is a fuzzy number, i.e. such fuzzy subset A of the real line \mathbb{R} , with membership function $\mu_{A_i} : \mathbb{R} \rightarrow [0, 1]$, that

- A_i is normal, i.e. there exist an element x_0 such that $\mu_{A_i}(x_0) = 1$,
- A_i is fuzzy convex, i.e. $\mu_{A_i}(\lambda x_1 + (1 - \lambda)x_2) \geq \mu_{A_i}(x_1) \wedge \mu_{A_i}(x_2)$, $(\forall x_1, x_2 \in \mathbb{R})$, $(\forall \lambda \in [0, 1])$,
- μ_{A_i} is upper semicontinuous,
- $supp(A_i)$ is bounded, where $supp(A_i) = cl(\{x \in \mathbb{R} : \mu_{A_i}(x) > 0\})$ and cl is the closure operator.

Let $A_\alpha = \{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$, $\alpha \in (0, 1]$, denote an α -cut of a fuzzy number A . Every α -cut of a fuzzy number is a closed interval, i.e. $A_\alpha = [A_L(\alpha), A_U(\alpha)]$, where $A_L(\alpha) = \inf\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$ and $A_U(\alpha) = \sup\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$.

A space of all fuzzy numbers will be denoted by $\mathbb{F}(\mathbb{R})$.

Although family $\mathbb{F}(\mathbb{R})$ is quite rich and consists of objects with diverse membership functions, fuzzy numbers with simpler membership functions are often preferred in practice. The most commonly used subclass of $\mathbb{F}(\mathbb{R})$ is formed by so-called *trapezoidal fuzzy numbers*, i.e. fuzzy numbers with linear sides. Thus a membership function of a trapezoidal fuzzy number is given by

$$\mu_T(x) = \begin{cases} 0 & \text{if } x < t_1, \\ \frac{x-t_1}{t_2-t_1} & \text{if } t_1 \leq x < t_2, \\ 1 & \text{if } t_2 \leq x \leq t_3, \\ \frac{t_4-x}{t_4-t_3} & \text{if } t_3 < x \leq t_4, \\ 0 & \text{if } x > t_4, \end{cases}$$

where $t_1 \leq t_2 \leq t_3 \leq t_4$. Since the membership function of a trapezoidal fuzzy number T is completely defined by these four real numbers we denote it usually as $T = T(t_1, t_2, t_3, t_4)$. If $t_2 = t_3$ we get the so-called *triangular fuzzy number*, while for $t_1 = t_2$ and $t_3 = t_4$ we get a *rectangular*

fuzzy number which is isomorphic with an interval. Finally, a trapezoidal fuzzy number such that $t_1 = t_2 = t_3 = t_4$, called *crisp number*, is isomorphic with a real number. Let us adopt the following notation: \mathbb{F}^T - a family of trapezoidal fuzzy numbers, \mathbb{F}^Δ - a family of triangular fuzzy numbers and \mathbb{P} - a family of intervals in \mathbb{R} (or rectangular fuzzy numbers).

The main goal of this paper is to find a median of a fuzzy sample A_1, \dots, A_n . However, when we move to fuzzy data and want to generalize this concept of the median to fuzzy context, we see immediately that it is not possible to make it straightforwardly since fuzzy numbers do not form a linear order and so we cannot find the “middle” one. Of course there are many methods for ranking fuzzy numbers, but there is too much subjectivity in them and hence no ranking method is generally accepted and broadly applied.

A fuzzy sample median $\widetilde{Med} = \widetilde{Med}(A_1, \dots, A_n)$ from the fuzzy random sample A_1, \dots, A_n was defined in [5] as a fuzzy number with α -cuts $\widetilde{Med}_\alpha(A_1, \dots, A_n) = [\widetilde{Med}_\alpha^L, \widetilde{Med}_\alpha^U]$ given by

$$\widetilde{Med}_\alpha^L = \begin{cases} (A_\alpha^L)_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{((A_\alpha^L)_{(\frac{n}{2})}) + ((A_\alpha^L)_{(\frac{n}{2}+1)})}{2} & \text{if } n \text{ is even,} \end{cases} \quad (1)$$

$$\widetilde{Med}_\alpha^U = \begin{cases} (A_\alpha^U)_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{((A_\alpha^U)_{(\frac{n}{2})}) + ((A_\alpha^U)_{(\frac{n}{2}+1)})}{2} & \text{if } n \text{ is even,} \end{cases} \quad (2)$$

where $(A_\alpha^L)_{(k)}$ denotes the k -th order statistic of a sample $(A_1)_L(\alpha), \dots, (A_n)_L(\alpha)$ while $(A_\alpha^U)_{(k)}$ is the k -th order statistic of a sample $(A_1)_U(\alpha), \dots, (A_n)_U(\alpha)$. It can be shown that the fuzzy sample median \widetilde{Med} becomes a traditional crisp sample median if the observations are not vague but crisp.

It was shown in [5] that the fuzzy sample median defined by (1) and (2) is a consistent estimator of a fuzzy population median. Moreover, some applications of so defined fuzzy sample median were given there, like a fuzzy sign test for verifying hypotheses on a population median based on fuzzy observations.

Unfortunately, the fuzzy sample median defined by (1) and (2) is very difficult to handle. It also does not have any straightforward interpretation and hence is not convenient for the practical use. Therefore, it would be nice to find another definition of a median of a fuzzy sample which assumes more regular shapes and hence is easier for further calculations and has more natural interpretation. To get a new idea let us turn back to classical statistics.

3. A fixed-shape fuzzy median

As it is known from basic probability theory, the expected value $E(X)$ is such a constant a which

minimizes $E(X - a)^2$. Similarly, the average is such a value y that minimizes $\sum_{i=1}^n (X_i - y)^2$, i.e.

$$\bar{X} = \arg \min_{y \in \mathbb{R}} \sum_{i=1}^n (X_i - y)^2.$$

On the other hand a probabilistic median minimizes $E|X - a|$ and hence a sample median can be perceived as a real value m which minimizes $\sum_{i=1}^n |X_i - y|$, i.e.

$$m = \arg \min_{y \in \mathbb{R}} \sum_{i=1}^n |X_i - y|. \quad (3)$$

It seems that (3) is a promising starting point for our further considerations.

Let us turn back to our fuzzy sample A_1, \dots, A_n and let d denote a distance between fuzzy numbers. Looking on (3) let us define a **fuzzy median** as such fuzzy number M which minimizes distance d between this number and a sample under study, i.e.

$$M = \arg \min_{B \in \mathbb{F}(\mathbb{R})} \sum_{i=1}^n d(B, A_i). \quad (4)$$

It is obvious that using different distances one may obtain different fuzzy medians. We will discuss this problem broadly later. Much more important is a remark that the minimization problem (4) seems to be too general. Actually, even if the solution of the minimization problem (4) exists, it may happen that the membership function of M is so complicated that it is very difficult to handle and hence it is useless in any further calculations. In such case we also face serious difficulties with its adequate interpretation. It is not only a pure academic problem. Indeed, calculations on fuzzy numbers described by not regular membership functions involve serious mathematical and numerical complications. Therefore, many researches have considered various approximations of fuzzy numbers to simplify future processing of these objects. It seems that the most popular is the trapezoidal approximation which substitutes any fuzzy number by a trapezoidal one which is in some sense close to the original one and possibly satisfy some other requirements (like invariance of some important parameters). The broad list of characteristics desired by the approximation of fuzzy numbers can be found in [7], while for the approximations algorithms we refer the reader to [1, 8, 16].

Trapezoidal approximation is usually satisfactory because it simplifies membership functions to linear and so eventually the result of approximation is represented by four real numbers only which determine its support and core. Simultaneously we get objects which remain fuzzy but at the same time are easier to explain. However, sometimes a greater simplification is still required and then the triangular or interval approximation might be recommended (see [6, 9]).

Keeping in mind all those efforts directed to simplification of the shapes of membership functions, it seems natural that some modifications of (4) would be desirable. In other words, we would like to consider a fuzzy median with a prescribed shape, i.e. such a fuzzy number which behaves like a median and simultaneously is simple enough. Now we are ready to introduce some useful definitions.

Definition 1 A *trapezoidal median* of a fuzzy sample A_1, \dots, A_n is such a trapezoidal fuzzy number $M^T \in \mathbb{F}^T$ which is closest to the fuzzy sample A_1, \dots, A_n with respect to distance d , i.e.

$$M^T = \arg \min_{B \in \mathbb{F}^T} \sum_{i=1}^n d(B, A_i). \quad (5)$$

Definition 2 A *triangular median* of a fuzzy sample A_1, \dots, A_n is such a triangular fuzzy number $M^\Delta \in \mathbb{F}^\Delta$ which is closest to the fuzzy sample A_1, \dots, A_n with respect to distance d , i.e.

$$M^\Delta = \arg \min_{B \in \mathbb{F}^\Delta} \sum_{i=1}^n d(B, A_i). \quad (6)$$

Definition 3 An *interval median* from a of sample A_1, \dots, A_n is such an interval $M^I \in \mathbb{P}$ which is closest to the fuzzy sample A_1, \dots, A_n with respect to distance d , i.e.

$$M^I = \arg \min_{B \in \mathbb{P}} \sum_{i=1}^n d(B, A_i). \quad (7)$$

Definition 4 A *crisp median* of a fuzzy sample A_1, \dots, A_n is such a real number $M^C \in \mathbb{R}$ which is closest to the fuzzy sample A_1, \dots, A_n with respect to distance d , i.e.

$$M^C = \arg \min_{B \in \mathbb{R}} \sum_{i=1}^n d(B, A_i). \quad (8)$$

Of course, we have here a kind of hierarchy, since each triangular fuzzy number is a trapezoidal fuzzy number and so on, but it seems that all four medians suggested above may be interesting in practice.

As we have mentioned earlier, using different metrics we may obtain various fuzzy medians. Hence in the next section we recall briefly the most important and the most popular distances utilized for fuzzy numbers. And later we will discuss the problem of the existence of fuzzy medians proposed in Def. 1-4.

Note, that a similar idea of a median but restricted to particular distances and interval data was discussed in [11, 13], while for fuzzy numbers and L^1 -type distance in [12]. On the other hand, a completely different idea of a median in the setting of imprecise probabilities was discussed in [3].

4. Metrics on the space of fuzzy numbers

A uniform type distance between fuzzy numbers is given by

$$D_U(A, B) = \max \left\{ \sup_{\alpha \in [0,1]} |A_L(\alpha) - B_L(\alpha)|, \sup_{\alpha \in [0,1]} |A_U(\alpha) - B_U(\alpha)| \right\}$$

Grzegorzewski [4] observed that for a fuzzy number A , the functions A_L and A_U are L_p -integrable. He introduced the metric $\delta_{p,q}$ given by

$$\delta_{p,q}^p(A, B) = (1-q) \int_0^1 |A_L(\alpha) - B_L(\alpha)|^p d\alpha + q \int_0^1 |A_U(\alpha) - B_U(\alpha)|^p d\alpha$$

where $1 \leq p < \infty$ and $0 < q < 1$. When $p = 2$ and $q = 1/2$ then using the notation $d = (1/2)^{-1/2} \delta_{p,q}$ we obtain the famous Euclidean distance given by

$$d(A, B)^2 = \int_0^1 (A_L(\alpha) - B_L(\alpha))^2 d\alpha + \int_0^1 (A_U(\alpha) - B_U(\alpha))^2 d\alpha \quad (9)$$

Zeng and Li [18] proposed a more general approach by introducing a nonnegative weight function $\lambda : [0, 1] \rightarrow \mathbb{R}$ such that $\int_0^1 \lambda(\alpha) d\alpha = 1/2$. The weighted distance d_λ is given by

$$d_\lambda(A, B)^2 = \int_0^1 (A_L(\alpha) - B_L(\alpha))^2 \lambda(\alpha) d\alpha + \int_0^1 (A_U(\alpha) - B_U(\alpha))^2 \lambda(\alpha) d\alpha.$$

More generally, Yeh [17] proposed the weighted L_2 -type distance d_λ ,

$$d_\lambda(A, B)^2 = \int_0^1 (A_L(\alpha) - B_L(\alpha))^2 \lambda_L(\alpha) d\alpha + \int_0^1 (A_U(\alpha) - B_U(\alpha))^2 \lambda_U(\alpha) d\alpha,$$

where $\lambda = (\lambda_L, \lambda_U)$, $\lambda_L, \lambda_U : [0, 1] \rightarrow \mathbb{R}$ are nonnegative weight functions such that $\int_0^1 \lambda_L(\alpha) d\alpha > 0$ and $\int_0^1 \lambda_U(\alpha) d\alpha > 0$. For $\lambda_L(\alpha) = \lambda_U(\alpha) = 1, \alpha \in [0, 1]$, we obtain the Euclidean metric.

Another class of distances between fuzzy numbers, introduced by Bertoluzza et al. [2], is given by

$$\tilde{D}_{f,\varphi}(A, B) = \left(\int_0^1 \tilde{D}_f^2(A_\alpha, B_\alpha) d\varphi(\alpha) \right)^{1/2},$$

such that

$$\begin{aligned} \tilde{D}_f^2([a, b], [c, d]) \\ = \int_0^1 (t|a - c| + (1-t)|b - d|)^2 df(t), \end{aligned}$$

where f is a normalized weight measure on $[0, 1]$ while function φ satisfies usually the following conditions: $\varphi(\alpha) \geq 0, \forall \alpha \in [0, 1]; \alpha_1 \leq \alpha_2 \Rightarrow \varphi(\alpha_1) \leq \varphi(\alpha_2)$ and $\int_0^1 \varphi(\alpha) d\alpha = 1$.

Expressing the above metric \tilde{D}_f in terms of the mid and spread of intervals, Trutschnig et al. [15] introduced the distance $D_{\psi,\theta}^*$ between fuzzy numbers as follows:

$$D_{\psi,\theta}^*(A, B) = \left(\int_0^1 (D_\theta^*(A_\alpha, B_\alpha))^2 d\psi(\alpha) \right)^{1/2},$$

where $\theta \in (0, 1], \psi$ is a weight probability measure on $[0, 1]$,

$$\begin{aligned} (D_\theta^*([a, b], [c, d]))^2 &= (\text{mid}[a, b] - \text{mid}[c, d])^2 \\ &\quad + \theta (\text{spr}[a, b] - \text{spr}[c, d])^2, \end{aligned}$$

and

$$\begin{aligned} \text{mid}[a_1, a_2] &= \frac{a_1 + a_2}{2} \\ \text{spr}[a_1, a_2] &= \frac{a_2 - a_1}{2}. \end{aligned}$$

5. The Rådström theorem

We will obtain the main results of the paper by using the following well-known embedding theorem of Rådström.

Theorem 5 ([10], Theorem 1) *Let us consider a triplet $(\mathbb{X}, +, \cdot)$ where $+: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X}$ and $\cdot : [0, \infty) \times \mathbb{X} \rightarrow \mathbb{X}$ satisfy the following properties:*

- (i) $(a + b) + c = a + (b + c)$ for all $a, b, c \in \mathbb{X}$;
- (ii) $a + b = b + a$ for all $a, b \in \mathbb{X}$;
- (iii) $a + c = b + c$ implies $a = b$ for all $a, b, c \in \mathbb{X}$;
- (iv) $\lambda(a + b) = \lambda a + \lambda b$ for all $\lambda \in [0, \infty)$ and $a, b \in \mathbb{X}$;
- (v) $(\lambda_1 + \lambda_2)a = \lambda_1 a + \lambda_2 a$ for all $\lambda_1, \lambda_2 \in [0, \infty)$ and $a \in \mathbb{X}$;
- (vi) $\lambda_1(\lambda_2 a) = \lambda_1 \lambda_2 a$ for all $\lambda_1, \lambda_2 \in [0, \infty)$ and $a \in \mathbb{X}$;
- (vii) $1a = a$ for all $a \in \mathbb{X}$.

Then there exists a vector space $(\widetilde{\mathbb{X}}, \oplus, \odot)$ and an injective application (inclusion) $i : \mathbb{X} \rightarrow \widetilde{\mathbb{X}}$ and, regarding \mathbb{X} as a subset of $\widetilde{\mathbb{X}}$ (that is adopting the convention $i(x) = x$ for all $x \in \mathbb{X}$) we have

$$\begin{aligned} a \oplus b &= a + b; \\ \lambda \odot a &= \lambda \cdot a \end{aligned}$$

for all $a, b \in \mathbb{X}$ and $\lambda \in [0, \infty)$.

If, in addition, there exists a metric d defined on \mathbb{X} satisfying

- (viii) $d(a + c, b + c) = d(a, b)$ for all $a, b, c \in \mathbb{X}$;
- (ix) $d(\lambda a, \lambda b) = \lambda d(a, b)$ for all $\lambda \in [0, \infty)$ and $a, b \in \mathbb{X}$;
- (x) $+: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X}$ and $\cdot : [0, \infty) \times \mathbb{X} \rightarrow \mathbb{X}$ are continuous on the topology generated by d on \mathbb{X} ,

then there exists a norm $\|\cdot\| : \widetilde{\mathbb{X}} \rightarrow [0, \infty)$ such that the metric \widetilde{d} on $\widetilde{\mathbb{X}}$ generated by $\|\cdot\|$ satisfies

$$d(a, b) = \widetilde{d}(a, b) \text{ for all } a, b \in \mathbb{X}.$$

Spaces like \mathbb{X} satisfying requirements (i)-(vii) in the above theorem are called semilinear spaces.

By Th. 5 we get easily

Corollary 6 *There exists a vector space $(\widetilde{\mathbb{F}(\mathbb{R})}, \oplus, \odot)$ so that*

$$A + B = A \oplus B \text{ for all } A, B \in \mathbb{F}(\mathbb{R})$$

and

$$\lambda A = \lambda \odot A \text{ for all } \lambda \in [0, \infty) \text{ and } A \in \mathbb{F}(\mathbb{R}).$$

In addition, if D is a metric on $\mathbb{F}(\mathbb{R})$ like those in the previous section, then there exists a metric \widetilde{D} which makes $(\widetilde{\mathbb{F}(\mathbb{R})}, \oplus, \odot, \widetilde{D})$ a normed space and such that

$$D(A, B) = \widetilde{D}(A, B) \text{ for all } A, B \in \mathbb{F}(\mathbb{R}).$$

6. Some existence results in approximation problems

By the following well known result (see, e.g., [14], Theorem 4.1.1)

Lemma 7 *Let (\mathbb{X}, d) be a metric space and let B be a compact subset of \mathbb{X} . Then for any $x \in \mathbb{X}$ there exists $x^* \in B$ such that*

$$d(x, x^*) = \inf_{y \in B} d(x, y).$$

we may prove the following lemma:

Lemma 8 *Let (\mathbb{X}, d) be a normed space and let Ω be a nonempty closed subset of a linear subspace of \mathbb{X} . Then, for any $x \in \mathbb{X}$ there exists $x^* \in \Omega$ such that*

$$d(x, x^*) = \inf_{y \in \Omega} d(x, y). \quad (10)$$

If, in addition, (\mathbb{X}, d) is a strictly convex normed space and Ω is convex then x^* is unique.

We also get the following result:

Theorem 9 *Let d be a metric defined on the space of fuzzy numbers $\mathbb{F}(\mathbb{R})$ which satisfies requirements (viii)-(x) of Th. 5 and let $(\widetilde{\mathbb{F}(\mathbb{R})}, \widetilde{d}, \oplus, \odot)$ be the normed space which realizes the embedding of $(\mathbb{F}(\mathbb{R}), d, +, \cdot)$ according to Th. 5. Let us consider a subset $\mathcal{A} \subseteq \mathbb{F}(\mathbb{R})$ for which there exists $\{v_2, v_3, \dots, v_m\} \subseteq \mathcal{A}$ such that:*

- (i) *The system $\{1, v_2, \dots, v_n\}$ is linearly independent in the vector space $(\widetilde{\mathbb{F}(\mathbb{R})}, \oplus, \odot)$;*
- (ii) $\mathcal{A} = \{\lambda_1 \cdot 1 + \sum_{i=2}^n \lambda_i v_i : \lambda_1 \in \mathbb{R}, \lambda_i \in [0, \infty), i \in \{2, \dots, n\}\}.$

Then \mathcal{A} is a closed subset of $\mathbb{F}(\mathbb{R})$ in the topology generated by the metric d .

Consequently, we obtain

Corollary 10 *Let d be a metric defined on the space of fuzzy numbers $\mathbb{F}(\mathbb{R})$ which satisfies requirements (viii)-(x) of Th. 5. Let us consider a subset $\mathcal{A} \subseteq \mathbb{F}(\mathbb{R})$ which satisfies requirements (i)-(ii) of Th. 9. If $(\widetilde{\mathbb{F}(\mathbb{R})}, \widetilde{d})$ is a normed space where \widetilde{d} is the extension of the metric d on $\widetilde{\mathbb{F}(\mathbb{R})}$ according to Th. 5 then \mathcal{A} is a closed subset of $\mathbb{F}(\mathbb{R})$ in the topology generated by \widetilde{d} on $\widetilde{\mathbb{F}(\mathbb{R})}$.*

Note that in general if (\mathbb{X}, δ) is a topological space and $\mathbb{X}_1 \subseteq \mathbb{X}$ and if A is a closed subset of \mathbb{X}_1 in the topology induced by δ on \mathbb{X}_1 , then it does not necessarily hold that A is a closed subset of \mathbb{X} in the topology δ . Therefore, the form of the set \mathcal{A} in Th. 9 is important in order to obtain the conclusion of the above corollary.

Combining Lemma 8 and Corollary 10, we easily obtain the following corollary.

Corollary 11 *Let d be a metric defined on $\mathbb{F}(\mathbb{R})$ which satisfies requirements (viii)-(x) of Th. 5 and let \mathcal{A} be a subset of $\mathbb{F}(\mathbb{R})$ like those from the hypothesis of Th. 9. Then for any $A \in \mathbb{F}(\mathbb{R})$ there exists $A^* \in \mathcal{A}$ such that*

$$d(A, A^*) = \inf_{B \in \mathcal{A}} d(A, B).$$

Now we may apply previous general results for some families of fuzzy numbers, especially important because of our interests in fuzzy medians defined in Def. 1-4.

Theorem 12 Let Ω be one of the following subsets of $\mathbb{F}(\mathbb{R})$: \mathbb{F}^T , \mathbb{F}^Δ , \mathbb{P} , \mathbb{R} . If d is a metric on $\mathbb{F}(\mathbb{R})$ which satisfies requirements (viii)-(x) of Th. 5 then Ω is a closed subset of $\mathbb{F}(\mathbb{R})$ in the topology generated by d . In addition, if $(\widetilde{\mathbb{F}(\mathbb{R})}, \tilde{d})$ is a normed space, where \tilde{d} is the extension of the metric d on $\widetilde{\mathbb{F}(\mathbb{R})}$ according to Th. 5, then Ω is a closed subset of $\widetilde{\mathbb{F}(\mathbb{R})}$ in the topology generated by \tilde{d} on $\widetilde{\mathbb{F}(\mathbb{R})}$.

Proof. The proof of the first statement is immediate since Ω has the properties of the set \mathcal{A} in the hypothesis of Th. 9. Suppose for example that $\Omega = \mathbb{F}^T$. Since any $T(t_1, t_2, t_3, t_4) \in \mathbb{F}^T$ can be written as

$$\begin{aligned} T &= t_1 \cdot 1 + (t_2 - t_1) v_1 \\ &\quad + (t_3 - t_2) v_3 + (t_4 - t_3) v_4, \end{aligned}$$

where $v_2 = (0, 1, 1, 1)$, $v_3 = (0, 0, 1, 1)$, $v_4 = (0, 0, 0, 1)$, clearly this implies that we can take $\mathcal{A} = \mathbb{F}^T$ in Th. 9. Similarly, we get the same conclusion for the other cases.

The proof of the second statement is immediate by Corollary 10 since by the above reasonings it results that Ω satisfies the hypothesis of this corollary too. ■

The following is the main result of this section.

Theorem 13 Let Ω be one of the following subsets of $\mathbb{F}(\mathbb{R})$: \mathbb{F}^T , \mathbb{F}^Δ , \mathbb{P} , \mathbb{R} . If d is a metric on $\mathbb{F}(\mathbb{R})$ satisfying requirements (viii)-(x) of Th. 5 then for any $A \in \Omega$ there exists $A^* \in \Omega$ such that

$$d(A, A^*) = \inf_{B \in \Omega} d(A, B).$$

Proof. By the proof of Th. 12 it results that Ω can be represented as the set \mathcal{A} from the hypothesis of Th. 9. Therefore, the desired conclusion easily follows from Corollary 11. ■

7. Existence of the fixed-shape fuzzy median

Let us consider a metric d defined on the space $\mathbb{F}(\mathbb{R})$ which satisfies requirements (viii)-(x) of Th. 5 and let $(\widetilde{\mathbb{F}(\mathbb{R})}, \tilde{d}, \oplus, \odot)$ be the normed space which realizes the embedding of $(\mathbb{F}(\mathbb{R}), d, +, \cdot)$. We construct now the power spaces of $\mathbb{F}(\mathbb{R})$ and $\widetilde{\mathbb{F}(\mathbb{R})}$ denoted with $\mathbb{F}(\mathbb{R})^n$ and $\widetilde{\mathbb{F}(\mathbb{R})}^n$, where $n \in \mathbb{N}$, $n \geq 2$ is fixed. By $A \in \mathbb{F}(\mathbb{R})^n$ we agree that $A = (A_1, \dots, A_n)$ where $A_i \in \mathbb{F}(\mathbb{R})$ for all $i \in \{1, \dots, n\}$. We adopt the same convention in the case of $\widetilde{\mathbb{F}(\mathbb{R})}$. Further one, we construct on $\widetilde{\mathbb{F}(\mathbb{R})}^n$ a vector space structure by extending the vector space structure of $\widetilde{\mathbb{F}(\mathbb{R})}$ using the well-known standard procedure. Then we notice that since $\mathbb{F}(\mathbb{R}) \subseteq \widetilde{\mathbb{F}(\mathbb{R})}$ we also have $\mathbb{F}(\mathbb{R})^n \subseteq \widetilde{\mathbb{F}(\mathbb{R})}^n$. We define metrics δ on $\mathbb{F}(\mathbb{R})^n$ and $\tilde{\delta}$ on $\widetilde{\mathbb{F}(\mathbb{R})}^n$, where

$$\delta(A, B) = \sum_{i=1}^n d(A_i, B_i)$$

and

$$\tilde{\delta}(A, B) = \sum_{i=1}^n \tilde{d}(A_i, B_i).$$

This means that for any $A, B \in \mathbb{F}(\mathbb{R})^n$ we have $\tilde{\delta}(A, B) = \delta(A, B)$. We are interested in a particular kind of set which will help us later to prove the existence of the fuzzy median. Suppose that \mathcal{A} is a subset of $\mathbb{F}(\mathbb{R})$ like those from the hypothesis of Th. 9. We introduce the diagonal set of \mathcal{A} in $\mathbb{F}(\mathbb{R})^n$, given by

$$D^n(\mathcal{A}) = \{(A, \dots, A) : A \in \mathcal{A}\}.$$

Since \mathcal{A} is closed in $\mathbb{F}(\mathbb{R})$ it easily follows that $D^n(\mathcal{A})$ is closed in $\mathbb{F}(\mathbb{R})^n$. Indeed, in order to prove this fact, let us consider in $D^n(\mathcal{A})$ the sequence $(\bar{A}_k)_{k \geq 1}$ such that $(\delta) \lim_{k \rightarrow \infty} \bar{A}_k = \bar{A}_0$. Denoting

$$\bar{A}_k = (A_k, \dots, A_k), \quad k \in \mathbb{N}, \quad k \geq 1$$

and

$$\bar{A}_0 = (A_0^1, A_0^2, \dots, A_0^n),$$

the convergence property and the definition of the metric δ easily implies that $(d) \lim_{k \rightarrow \infty} \bar{A}_k = A_0^i$, $i \in \{1, \dots, n\}$ and, since \mathcal{A} is closed in $\mathbb{F}(\mathbb{R})$ we get that $A_0^i \in \mathcal{A}$ for all $i \in \{1, \dots, n\}$. The uniqueness of the limit implies $A_0^1 = \dots = A_0^n$ and we thus obtain that $\bar{A}_0 \in D^n(\mathcal{A})$. Clearly, this implies that $D^n(\mathcal{A})$ is closed in $\mathbb{F}(\mathbb{R})^n$. Analogously, if \mathcal{A} is a closed subset of $\widetilde{\mathbb{F}(\mathbb{R})}$ then we get that $D^n(\mathcal{A})$ is a closed subset of $\widetilde{\mathbb{F}(\mathbb{R})}^n$ too.

We have the following main result of this section.

Theorem 14 Let d be a metric defined on $\mathbb{F}(\mathbb{R})$ which satisfies requirements (viii)-(x) of Th. 5. Let us consider an arbitrary sample of fuzzy numbers, A_1, \dots, A_n and let Ω be one of the following subsets of $\mathbb{F}(\mathbb{R})$: \mathbb{F}^T , \mathbb{F}^Δ , \mathbb{P} , \mathbb{R} . Then there exists a fuzzy median of the sample with respect to Ω and the metric d .

Proof. By the proof of Th. 12 it results that Ω satisfies the hypothesis in Th. 9 and Corollary 10 by taking $\mathcal{A} = \Omega$ there. Analyzing the form of \mathcal{A} (and hence Ω) in Th. 9, it is easily seen that Ω is a closed subset of a finite dimensional linear subspace $\mathcal{L} \subseteq \widetilde{\mathbb{F}(\mathbb{R})}$. Now, since d satisfies requirements (viii)-(x) of Th. 5, let $(\widetilde{\mathbb{F}(\mathbb{R})}, \tilde{d}, \oplus, \odot)$ be the normed space which realizes the embedding of $(\mathbb{F}(\mathbb{R}), d, +, \cdot)$. This easily implies that the power space $(\widetilde{\mathbb{F}(\mathbb{R})}^n, \tilde{\delta}, \oplus, \odot)$ is a normed space too. Taking into account the properties of Ω we obtain that $D^n(\Omega)$ is a closed convex subset of the linear subspace $\mathcal{L}^n \subseteq \widetilde{\mathbb{F}(\mathbb{R})}^n$. Taking $A = (A_1, \dots, A_n)$, by Lemma 8, it results the existence of $\bar{A}_0 \in D^n(\Omega)$, $\bar{A}_0 = (A_0, \dots, A_0)$, such that

$$\tilde{\delta}(A, \bar{A}_0) \leq \tilde{\delta}(A, \bar{B}) \quad (11)$$

for all $\bar{B} \in D^n(\Omega)$. Now, let us choose arbitrary $B \in \Omega$. We observe that $\bar{B} \in D^n(\Omega)$ where $\bar{B} = (B, \dots, B)$. Taking into account the properties of \tilde{d} and \tilde{d} we get

$$\sum_{i=0}^n d(A_i, A_0) = \tilde{d}(A, \bar{A}_0)$$

and

$$\sum_{i=0}^n d(A_i, B) = \tilde{d}(A, \bar{B}),$$

which by relation (11) implies

$$\sum_{i=0}^n d(A_i, A_0) \leq \sum_{i=0}^n d(A_i, B).$$

This means that A_0 is a fuzzy median of the sample A_1, \dots, A_n with respect to Ω and the metric d . ■

Note that in the above theorem the convexity of Ω is not needed for the existence result because in Lemma 8 the convexity is not required. But if the product space $(\widetilde{\mathbb{F}(\mathbb{R})}^n, \tilde{d})$ is strictly convex then the fuzzy median is unique with respect to Ω because it is very easy to prove that Ω is a convex subset of $\widetilde{\mathbb{F}(\mathbb{R})}$. At the end of this section let us remark that all the distances or families of distances between fuzzy numbers, exemplified in Sec. 4, satisfy the hypothesis in Th. 14. Hence the fixed-shape fuzzy median always exists with respect to those distances.

8. Conclusions

Problems with ordering of fuzzy numbers is the reason that when trying to determine the median of a fuzzy sample we should apply the minimum distance method rather than looking for the center of the ordered sample. Moreover, a care for the ease of further calculations and data processing induce to searching objects that are relatively simple. These two arguments founds the basis for our approach for defining a fixed-shape fuzzy median of a fuzzy sample. The theorems given in the paper guarantee the existence of such a fuzzy median for the most important subfamilies of fuzzy numbers and the broad class of metrics applied in real-life problems.

For the future it would be interesting to study the fixed-shape fuzzy median by classifying the metrics into metrics generated by inner products, metrics generated by norms from strictly (or even uniformly) convex normed spaces and metrics generated from norms which are not in the previous categories. In this way we can find further interpretation for the fixed-shape fuzzy median. For example it can be proved whether the fixed-shape fuzzy median is unique or not.

Acknowledgements

The work of the first and second author was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-ID-PCE-2011-3-0861.

References

- [1] A. I. Ban, Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval, *Fuzzy Sets and Systems*, 159:1327-1344, 2008.
- [2] C. Bertoluzza, N. Corral and A. Salas, On a new class of distances between fuzzy numbers, *Mathware and Soft Computing*, 2:71-84, 1995.
- [3] I. Couso and L. Sánchez, The behavioral meaning of the median, In: C. Borgelt et al. (Eds.), *Combining Soft Computing and Statistical Methods in Data Analysis*, Springer, Heidelberg, pp. 115-123, 2010.
- [4] P. Grzegorzewski, Metrics and orders in space of fuzzy numbers, *Fuzzy Sets and Systems*, 97:83-94, 1998.
- [5] P. Grzegorzewski, Statistical inference about the median from vague data, *Control and Cybernetics*, 27:447-464, 1998.
- [6] P. Grzegorzewski, Nearest interval approximation of a fuzzy number, *Fuzzy Sets and Systems*, 130:321-330, 2002.
- [7] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers, *Fuzzy Sets and Systems*, 153:115-135, 2005.
- [8] P. Grzegorzewski, Trapezoidal approximations of fuzzy numbers preserving the expected interval-Algorithms and properties, *Fuzzy Sets and Systems*, 47:1354-1364, 2008.
- [9] P. Grzegorzewski, On the interval approximation of fuzzy numbers, *Communications in Computer and Information Science*, 299: 58-68, 2012.
- [10] H. Rådström, An embedding theorem for spaces of convex sets, *Proceedings of the American Mathematical Society*, 3:165-169, 1952.
- [11] B. Sinova, M.R. Casals, A. Colubi and M.A. Gil, The median of a random interval, In: C. Borgelt et al. (Eds.), *Combining Soft Computing and Statistical Methods in Data Analysis*, Springer, Heidelberg, pp. 575-583, 2010.
- [12] B. Sinova, M.A. Gil, A. Colubi and S. Van Aelst, The median of a random fuzzy number. The 1-norm distance approach, *Fuzzy Sets and Systems*, 200:99-115, 2012.
- [13] B. Sinova, G. Gonzales-Rodriguez and S. Van Aelst, An alternative approach to the median of a random interval using an L^2 metric, In: R. Kruse et al. (Eds.), *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, Springer, Heidelberg, pp. 273-281, 2013.

- [14] D. D. Stancu, G. Coman, O. Agratini and R. Trâmbițaș, *Analiză numerică și teoria aproximării* (in Romanian), Vol I, Presa Universitară Clujeană, Cluj-Napoca, 2001.
- [15] W. Trutschnig, G. González-Rodríguez, A. Colubi and M. A. Gil, A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread, *Information Sciences*, 179:3964-3972, 2009.
- [16] C-T. Yeh, Trapezoidal and triangular approximations preserving the expected interval, *Fuzzy Sets and Systems*, 159:1345-1353, 2008.
- [17] C-T. Yeh, Weighted trapezoidal and triangular approximations of fuzzy numbers, *Fuzzy Sets and Systems*, 160:3059-3079, 2009.
- [18] W. Zeng and H. Li, Weighted triangular approximation of fuzzy numbers, *International Journal of Approximate Reasoning*, 46:137-150, 2007.