ON THE ABSOLUTE VALUE OF TRAPEZOIDAL FUZZY NUMBERS AND THE MANHATTAN DISTANCE OF FUZZY VECTORS

Julio Rojas-Mora
UMR ESPACE 6012 CNRS, Université d'Avignon (UAPV), Avignon, France

Jaime Gil-Lafuente
Dept. of Business Economics and Organization, University of Barcelona, Barcelona, Spain

Didier Josselin
UMR ESPACE 6012 CNRS, Université d'Avignon (UAPV), Avignon, France

Keywords: Fuzzy sets, Manhattan distance, Absolute value.

Abstract: The computation of the Manhattan distance for fuzzy vectors composed of trapezoidal fuzzy numbers (TrFN) requires the application of the absolute value to the differences between components. The membership function of the absolute value of a fuzzy number has been defined by Dubois and Prade as well as by Chen and Wang. The first one only removes the negative values of the fuzzy number, increasing its expected value. Conversely, Chen and Wang's definition maintains the expected value, but can produce a TrFN with negative values. In this paper, we present the "positive correction" of the absolute value, a method to remove the negative values of a TrFN while maintaining its expected value. This operation also complies with a logic principle of any uncertain distance: reducing the distance should also reduce its uncertainty.

1 INTRODUCTION

In several fields, the necessity to determine the distance that separates two points in $\mathbb{R}^n$ arises. When there is uncertainty on the location of these points, the calculation of the distance has to take this uncertainty into consideration. By modeling uncertainty with fuzzy subsets (Zadeh, 1965), it is possible to calculate some form of distance that complies with this consideration.

The literature is broad in this area, but a non-comprehensive list of publications has to include the work of (Voxman, 1998), who calculated crisp metrics between two fuzzy numbers, but who also questioned this approach, studying fuzzy distances between them. (Tran and Duckstein, 2002), in the context of fuzzy numbers' ranking, proposed a distance function that takes into account all points in the fuzzy numbers compared. (Chen and Wang, 2008) defined a fuzzy distance that uses the absolute value of a fuzzy number, calculated through its graded mean integration representation (GMIR). Finally, (Li and Liu, 2008) make use of an expected value operator to define a metric space of fuzzy variables.

In this paper, we will go back to the simplest representation of the distance between two points, the Manhattan distance. Applying this distance to trapezoidal fuzzy numbers (TrFN), we would like to obtain a fuzzy number as a result, reflecting the uncertainty on the distance itself. Nonetheless, we will subject this distance to some conditions. Firstly, when the distance is reduced, so must do its uncertainty. By this, we mean that the uncertainty we have while assessing a distance of about 20 Km has to be much bigger that the uncertainty assessing a distance of 5 cm.

Secondly, the distance has to be positive at all times, as a negative distance has no sense in the real world. Finally, because we will be operating with TrFN, we would like that the distance is also a TrFN.

For the calculation of the Manhattan distance between two fuzzy numbers, there is the need to define the absolute value of a fuzzy number. We will explore the approaches followed by (Dubois and Prade,

Rojas-Mora J., Gil-Lafuente J. and Josselin D..
ON THE ABSOLUTE VALUE OF TRAPEZOIDAL FUZZY NUMBERS AND THE MANHATTAN DISTANCE OF FUZZY VECTORS.
DOI: 10.5220/0003674203990406
In *Proceedings of the International Conference on Evolutionary Computation Theory and Applications* (FCTA-2011), pages 399-406
ISBN: 978-989-8425-83-6

1979) and (Chen and Wang, 2008), analyzing their shortcomings according to the conditions previously posed and proposing a solution that takes them into consideration.

This paper introduces the basic concepts of fuzzy sets and fuzzy numbers in Section 2. Then, on Section 3, we will study the Manhattan distance as well as the definitions of the absolute value for fuzzy numbers previously cited. Section 4 explains the procedure needed to overcome the problems that arise from the use of the absolute value according to our constraints. Finally, on Section 5 we will present the conclusions of this work.

2 FUZZY SUBSETS AND FUZZY NUMBERS

In this section, we will introduce the basic definitions use throughout this paper. We will start by defining what a fuzzy subset is and what properties we will ask of them.

Definition 1. *A fuzzy subset $\tilde{F}$ is a set whose elements may not follow the law of excluded middle that rules over Boolean logic, i.e., their membership function can be mapped as:*

$$\mu_{\tilde{F}} : X \to [0,1].$$

In general, a fuzzy subset $\tilde{F}$ can be represented by a set of pairs consistent of the elements x of the universal set X and a grade of membership $\mu_{\tilde{A}}(x)$:

$$\tilde{F} = \{(x, \mu_{\tilde{F}}(x)) \mid x \in X, \mu_{\tilde{F}}(x) \in [0,1]\}. \quad (1)$$

Definition 2. *An α-cut of a fuzzy subset $\tilde{F}$ is defined by:*

$$F_\alpha = \{x \in X : \mu_{\tilde{F}}(x) \geq \alpha\}, \quad (2)$$

i.e., the subset of all elements that belong to $\tilde{F}$ at least in a degree α.

Definition 3. *A fuzzy subset $\tilde{F}$ is convex, if and only if:*

$$\lambda x_1 + (1 - \lambda x_2) \in F_\alpha \, \forall x_1, x_2 \in F_\alpha, \, \alpha, \lambda \in [0,1], \quad (3)$$

i.e., all the points in $[x_1, x_2]$ must belong to A_α, for any α.

Definition 4. *A fuzzy subset is "normal" if and only if*

$$\max_{x \in X} (\mu_{\tilde{F}}(x)) = 1. \quad (4)$$

Definition 5. *The "core" of a normal fuzzy subset is:*

$$N_{\tilde{F}} = \{x : \mu_{\tilde{F}}(x) = 1\}. \quad (5)$$

Now, we will define a fuzzy number based on these definitions.

Definition 6. *A fuzzy number $\underset{\sim}{M}$ is a convex, normal fuzzy subset with domain in $\mathbb{R}$, for which:*

1. $\bar{x} := N_{\underset{\sim}{M}}$, $\mathrm{card}(\bar{x}) = 1$, and
2. $\mu_{\underset{\sim}{M}}$ is, at least, piecewise continuous.

The first condition is usually dropped as there is a tendency, that we will follow, to call "fuzzy numbers" those fuzzy subsets for which the core has more than one element (Zimmermann, 2005, 57).

Definition 7. *A TrFN is defined by the membership function:*

$$\mu_{\underset{\sim}{M}}(x) = \begin{cases} 1 - \frac{x_2 - x}{x_2 - x_1}, & \text{if } x_1 \leq x < x_2 \\ 1, & \text{if } x_2 \leq x \leq x_3 \\ 1 - \frac{x - x_3}{x_4 - x_3}, & \text{if } x_3 < x \leq x_4 \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

A TrFN is represented by a 4-tuple whose first and fourth elements correspond to the extremes from where the membership function begins to grow, and whose second and third components are the limits of the interval where the maximum certainty lies, i.e., $\underset{\sim}{M} = (x_1, x_2, x_3, x_4)$. From this point on, and for easiness while operating with several TrFN's, we will change the "x" in the 4-tuple for the lowercase letter that names a TrFN, i.e., $\underset{\sim}{M} = (m_1, m_2, m_3, m_4)$.

It can be complicated to compare fuzzy numbers with the naked eye due to their uncertain nature, being necessary to remove all entropy in a process called defuzzification. There are several methods to do it, but we selected the graded mean integration representation (GMIR) as stated by (Chen and Hsieh, 1999).

Definition 8. *The GMIR of a non-normal TrFN is:*

$$E\left(\underset{\sim}{M}\right) = \frac{\int_0^{\max\left(\mu_{\underset{\sim}{M}}\right)} \frac{\mu}{2}\left(L^{-1}_{\underset{\sim}{M}}(\mu) + R^{-1}_{\underset{\sim}{M}}(\mu)\right) d\mu}{\int_0^{\max\left(\mu_{\underset{\sim}{M}}\right)} \mu d\mu}. \quad (7)$$

where $L^{-1}(\mu)$ and $R^{-1}(\mu)$ are the inverse functions that define the TrFN in $[m_1, m_2]$ and $[m_3, m_4]$, respectively.

Remark 1. *For a TrFN, the GMIR is:*

$$E\left(\underset{\sim}{M}\right) = \frac{m_1 + 2m_2 + 2m_3 + m_4}{6}. \quad (8)$$

We can see the GMIR of a TrFN as a weighted mean value where central components have double the weight of the exterior ones. It is also a form of expected value of the TrFN.

Fuzzy numbers arithmetic is derived from Zadeh's extension principle (Zadeh, 1975). It can also be defined through interval arithmetic for every α−cut, as done by (Kaufmann and Gupta, 1985).

Definition 9. *The addition of two TrFN's* $\underset{\sim}{M}$ *and* $\underset{\sim}{N}$ *is defined as:*

$$\underset{\sim}{M} \oplus \underset{\sim}{N} = (m_1+n_1, m_2+n_2, m_3+n_3, m_4+n_4). \quad (9)$$

Definition 10. *The "image" of a TrFN* $\underset{\sim}{M}$ *is:*

$$\underset{\sim}{M}^- = (-m_4, -m_3, -m_2, -m_1). \quad (10)$$

Definition 11. *The subtraction of two TrFN's* $\underset{\sim}{M}$ *and* $\underset{\sim}{N}$ *is defined as:*

$$\begin{aligned} \underset{\sim}{M} \ominus \underset{\sim}{N} &= \underset{\sim}{M} \oplus \underset{\sim}{N}^- \\ &= (m_1-n_4, m_2-n_3, m_3-n_2, m_4-n_1). \end{aligned} \quad (11)$$

3 MANHATTAN DISTANCE

The Manhattan or L_1 distance, is the simplest separation measurement between two vectors in an $n-$dimensional space.

Definition 12. *For two fuzzy vectors* $\underset{\sim}{A}$ *and* $\underset{\sim}{B}$*, the Manhattan distance is defined by:*

$$d_H\left(\underset{\sim}{A}, \underset{\sim}{B}\right) = \sum_{i=1}^{n} \left| \underset{\sim}{A}^{(i)} \ominus \underset{\sim}{B}^{(i)} \right|. \quad (12)$$

In this definition we have to deal with the absolute value of a TrFN. We will see two different ways in which its characteristic function has been defined in the literature. The first one, by (Dubois and Prade, 1979), uses Zadeh's extension principle.

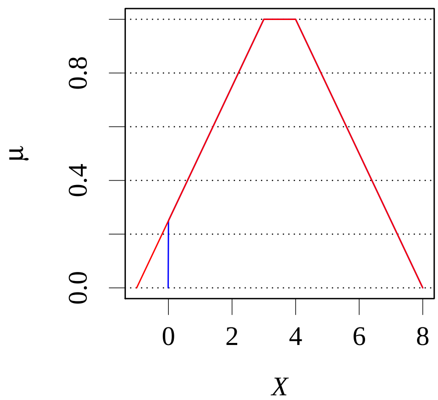
Definition 13 ((Dubois and Prade, 1979))**.** *The absolute value of a fuzzy number is defined as :*

$$\mu_{\left|\underset{\sim}{A}\right|}(x) = \begin{cases} \max\left(\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{A}}(-x)\right) & , \textit{if } x \geq 0 \\ 0 & , \textit{else.} \end{cases} \quad (13)$$

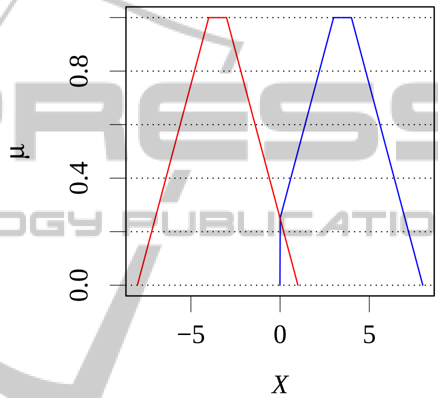
For example, if $\underset{\sim}{A} = (3,5,6,9)$ and $\underset{\sim}{B} = (1,2,2,4)$ we can see in Figure 1 that by using (13) in (12), the negative side of the TrFN is truncated and the distance between two TrFN is not a TrFN. We will now prove that by this truncation the GMIR of the distance between two TrFN's is bigger than the absolute value of the GMIR of their difference.

Proposition 1. *By Definition 13,* $\left|E\left(\underset{\sim}{A} \ominus \underset{\sim}{B}\right)\right| \leq E\left(\left|\underset{\sim}{A} \ominus \underset{\sim}{B}\right|\right)$.

Proof. Let $\underset{\sim}{C} = (c_1, c_2, c_3, c_4) = \underset{\sim}{A} \ominus \underset{\sim}{B}$. If $c_1 \geq 0$, then $\mu_{\left|\underset{\sim}{C}\right|}(x) = \mu_{\underset{\sim}{C}}(x), \forall x$ and $E\left(\left|\underset{\sim}{C}\right|\right) = \left|E\left(\underset{\sim}{C}\right)\right|$. Conversely, if $c_4 \leq 0$, then $\mu_{\left|\underset{\sim}{C}\right|}(x) = \mu_{\underset{\sim}{C}}(-x), \forall x$. Thus:



a) In red $\underset{\sim}{A} \ominus \underset{\sim}{B}$ and in blue $|\underset{\sim}{A} \ominus \underset{\sim}{B}|$.



b) In red $\underset{\sim}{B} \ominus \underset{\sim}{A}$ and in blue $|\underset{\sim}{B} \ominus \underset{\sim}{A}|$.

Figure 1: Manhattan distance between two TrFN.

$$\begin{aligned} E\left(\left|\underset{\sim}{C}\right|\right) &= \left|E\left(\underset{\sim}{C}\right)\right| \\ \frac{-c_4-2c_3-2c_2-c_1}{6} &= \left|\frac{c_1+2c_2+2c_3+c_4}{6}\right| \\ \frac{-c_4-2c_3-2c_2-c_1}{6} &= -\left(\frac{c_1+2c_2+2c_3+c_4}{6}\right) \\ \frac{-c_4-2c_3-2c_2-c_1}{6} &= \frac{-c_1-2c_2-2c_3-c_4}{6} \end{aligned}$$

Now, if either $c_1 < 0$ and $c_i > 0, \forall i \in \{2,3,4\}$, or $c_i < 0, \forall i \in \{1,2,3\}$ and $c_4 > 0$, then the shape of $\left|\underset{\sim}{C}\right|$ is that of the blue fuzzy number in Figure 1.b. Without losing any generality let us assume in this part of the proof that we only have the first case, i.e., $c_1 < 0$ and $c_i > 0, \forall i \in \{2,3,4\}$. Let us denote by $L_{\left|\underset{\sim}{C}\right|}(x)$, respectively $R_{\left|\underset{\sim}{C}\right|}(x)$, the membership function that defines $\left|\underset{\sim}{C}\right|$ in $[0, c_2]$, respectively in $[c_3, c_4]$. Then:

$$
\begin{aligned}
L_{\left|\underset{\sim}{C}\right|}(x) &= \frac{x-c_1}{c_2-c_1} &(14)\\
R_{\left|\underset{\sim}{C}\right|}(x) &= 1-\frac{x-c_3}{c_4-c_3}. &(15)
\end{aligned}
$$

Let $\mu_{12} = \mu_{\underset{\sim}{C}}(0)$. By (15):

$$\mu_{12} = -\frac{c_1}{c_2-c_1}. \tag{16}$$

Using (16) as breaking point, we will find $E\left(\left|\underset{\sim}{C}\right|\right)$:

$$
\begin{aligned}
E\left(\left|\underset{\sim}{C}\right|\right) &= \frac{\int_0^1 \frac{\mu}{2}\left(L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)+R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)\mathrm{d}\mu}{\int_0^1 \mu\,\mathrm{d}\mu}\\
&= \frac{\int_0^{\mu_{12}} \frac{\mu}{2}\left(L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)+R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)\mathrm{d}\mu}{\int_0^1 \mu\,\mathrm{d}\mu} + \\
&\quad \frac{\int_{\mu_{12}}^1 \frac{\mu}{2}\left(L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)+R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)\mathrm{d}\mu}{\int_0^1 \mu\,\mathrm{d}\mu}
\end{aligned}
$$

By (13) , $L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu) = 0, \mu \in [0,\mu_{12}]$. Thus:

$$
\begin{aligned}
E\left(\left|\underset{\sim}{C}\right|\right) &= \frac{\int_0^{\mu_{12}} \frac{\mu}{2}\left(R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)\mathrm{d}\mu}{\int_0^1 \mu\,\mathrm{d}\mu} + \\
&\quad \frac{\int_{\mu_{12}}^1 \frac{\mu}{2}\left(L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)+R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)}{\int_0^1 \mu\,\mathrm{d}\mu}\mathrm{d}\mu\\
&= \frac{c_1+2c_2+2c_3+c_4}{6} - \\
&\quad \frac{{c_1}^3}{6\left(c_1-c_2\right)^2}.
\end{aligned}
$$

Now, we state the proposition:

$$
\begin{aligned}
\left|E\left(\underset{\sim}{C}\right)\right| &\le E\left(\left|\underset{\sim}{C}\right|\right)\\
\tfrac{c_1+2c_2+2c_3+c_4}{6} &\le \tfrac{c_1+2c_2+2c_3+c_4}{6} - \tfrac{{c_1}^3}{6\left(c_1-c_2\right)^2}\\
\frac{{c_1}^3}{\left(c_1-c_2\right)^2} &\le 0. &(17)
\end{aligned}
$$

The denominator of (17) is always positive, while the numerator is always negative, thus, proving the proposition. In the final case, $c_1 \le c_2 < 0$ and $0 < c_3 \le c_4$. If $E(\underset{\sim}{C}) > 0$ then $\left|\underset{\sim}{C}\right| = (0,0,c_3,c_4)$, else $\left|\underset{\sim}{C}\right| = (0,0,-c_2,-c_1)$. Again, without any lack of generality, we will prove the proposition for the first case only:

$$E\left(\left|\underset{\sim}{C}\right|\right) = \frac{\int_0^1 \frac{\mu}{2}\left(L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)+R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)\mathrm{d}\mu}{\int_0^1 \mu\,\mathrm{d}\mu}.$$

For this case $L^{-1}_{\left|\underset{\sim}{C}\right|}(\mu) = 0, \mu \in [0,1]$, thus:

$$
\begin{aligned}
E\left(\left|\underset{\sim}{C}\right|\right) &= \frac{\int_0^1 \frac{\mu}{2}\left(R^{-1}_{\left|\underset{\sim}{C}\right|}(\mu)\right)\mathrm{d}\mu}{\int_0^1 \mu\,\mathrm{d}\mu}\\
&= \frac{2c_3+c_4}{6}.
\end{aligned}
$$

Stating the proposition:

$$
\begin{aligned}
\left|E\left(\underset{\sim}{C}\right)\right| &\le E\left(\left|\underset{\sim}{C}\right|\right)\\
\frac{c_1+2c_2+2c_3+c_4}{6} &\le \frac{2c_3+c_4}{6}\\
\frac{c_1+2c_2}{6} &\le 0. &(18)
\end{aligned}
$$

As both c_1 and c_2 are negative, the proposition is proved. $\square$

So, not only using Dubois and Prade's definition makes that the absolute value of a TrFN is not always a TrFN, but it also overestimates the expected value obtained by (8). There is a third problem that we will see now.

Proposition 2. *By Definition 13,* $E\left(\left|\underset{\sim}{A} \ominus \underset{\sim}{A}\right|\right) \neq 0$.

Proof. By Definition 11:

$\underset{\sim}{A} \ominus \underset{\sim}{A} = \underset{\sim}{A} \oplus \underset{\sim}{A}^- = (a_1-a_4, a_2-a_3, a_3-a_2, a_4-a_1)$,

which is a symmetric, zero centered TrFN. By Definition 13, the absolute value of a symmetric, zero centered TrFN is a TrFN that only covers the right side of the first one, i.e., $\left|\underset{\sim}{A} \ominus \underset{\sim}{A}\right| = (0,0,a_3-a_2,a_4-a_1)$, thus $E\left(\left|\underset{\sim}{A} \ominus \underset{\sim}{A}\right|\right) \neq 0$. $\square$

A second definition of the absolute value of a TrFN is directly based on the GMIR, defuzzifying the fuzzy number to determine whether it is positive or not.

Definition 14. *Given a TrFN* $\underset{\sim}{M}$*, by means of* (8)*:*

$$
\begin{aligned}
&\underset{\sim}{M} < 0, \quad \textit{if } E(\underset{\sim}{M}) < 0,\\
&\underset{\sim}{M} = 0, \quad \textit{if } E(\underset{\sim}{M}) = 0,\\
&\underset{\sim}{M} > 0, \quad \textit{if } E(\underset{\sim}{M}) > 0.
\end{aligned}
$$

Definition 15 ((Chen and Wang, 2008))**.** *The absolute value of a TrFN* $\underset{\sim}{M}$ *is:*

$$\left|\underset{\sim}{M}\right| = \begin{cases} \underset{\sim}{M}, & \textit{if } \underset{\sim}{M} > 0,\\ 0, & \textit{if } \underset{\sim}{M} = 0,\\ \underset{\sim}{M}^-, & \textit{if } \underset{\sim}{M} < 0. \end{cases} \tag{19}$$

This definition of the absolute value has two advantages over that of (Dubois and Prade, 1979). In first place, $\left|\underset{\sim}{A} \ominus \underset{\sim}{B}\right|$ is always a TrFN, and in second place, $E\left(\left|\underset{\sim}{A} \ominus \underset{\sim}{A}\right|\right)=0$, as $\underset{\sim}{A} \ominus \underset{\sim}{A}$ is a symmetric, zero centered TrFN.

Nonetheless, it has a problem of its own. As there is no transformation in shape, the absolute value of a TrFN might have negative values. For example, if $\underset{\sim}{C}=(-1,2,4,6)$, then $E\left(\underset{\sim}{C}\right)=\frac{-1+2\cdot 2+2\cdot 4+6}{6}=2.833>0$, so $\left|\underset{\sim}{C}\right|=\underset{\sim}{C}=(-1,2,4,6)$. This result is particularly problematic for a distance, that by definition cannot be negative.

There is a second problem with the definition of (Chen and Wang, 2008) when we use it for the Manhattan distance. As the distance goes to zero, so must do it the uncertainty that about it we have, which is modeled by the area covered by the TrFN. In mathematical terms, this means that:

$$\lim_{E\left(\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)\right)\to 0} \int \underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right) \mathrm{d}x = 0. \tag{20}$$

In the following proposition, we will see this is not the behavior of the Manhattan distance based on Definition 15.

Proposition 3. *By Definitions 12 and 15,* $\exists \underset{\sim}{A},\underset{\sim}{B}$: $\lim_{E\left(\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)\right)\to 0} \int \underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right) dx \neq 0$.

Proof. From (12) and (19):

$$\begin{aligned} \underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)=0 &\iff E\left(\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)\right)=0 \\ E\left(\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)\right)=0 &\iff E\left(\left|\underset{\sim}{A}\ominus\underset{\sim}{B}\right|\right)=0 . \end{aligned}$$

Let's define the zero centered, symmetric TrFN $\underset{\sim}{C}=\underset{\sim}{A}\ominus\underset{\sim}{B}=(c_1,c_2,c_3,c_4)=(-c_4,-c_3,c_3,c_4)$, that is:

$$\begin{aligned} -c_4 &= a_1-b_4=-a_4+b_1 \\ -c_3 &= a_2-b_3=-a_3+b_2 \\ c_3 &= a_3-b_2 \\ c_4 &= a_4-b_1 . \end{aligned}$$

Now, by (8):

$$\begin{aligned} E\left(\underset{\sim}{C}\right) &= \frac{c_1+2\cdot c_2+2\cdot c_3+c_4}{6} \\ &= \frac{-c_4-2\cdot c_3+2\cdot c_3+2\cdot c_4}{6} \\ &= 0, \end{aligned}$$

thus $\left|\underset{\sim}{C}\right|=0$ and $\int\left|\underset{\sim}{C}\right| \mathrm{d}x=0$. Now, if a scalar ε is added to $\underset{\sim}{A}$, then:

$$\underset{\sim}{A}'=\underset{\sim}{A}+\varepsilon=(a_1+\varepsilon,a_2+\varepsilon,a_3+\varepsilon,a_4+\varepsilon) ,$$

so $\underset{\sim}{C}'=\underset{\sim}{A}'\ominus\underset{\sim}{B}=(c'_1,c'_2,c'_3,c'_4)$ and:

$$\begin{aligned} c'_1 &= a'_1-b'_4=a_1-b_4+\varepsilon=c_1+\varepsilon \\ c'_2 &= a'_2-b'_3=a_2-b_3+\varepsilon=c_2+\varepsilon \\ c'_3 &= a'_3-b'_2=a_3-b_2+\varepsilon=c_3+\varepsilon \\ c'_4 &= a'_4-b'_1=a_4-b_1+\varepsilon=c_4+\varepsilon . \end{aligned}$$

Then:

$$\begin{aligned} E\left(\underset{\sim}{C}'\right) &= \frac{c'_1+2\cdot c'_2+2\cdot c'_3+c'_4}{6} \\ &= \frac{1}{6}\left((-a_4+b_1+\varepsilon)+2(-a_3+b_2+\varepsilon)+\right. \\ &\quad \left.2(a_3-b_2+\varepsilon)+(a_4-b_1+\varepsilon)\right) \\ &= \varepsilon \end{aligned}$$

thus, $\forall\varepsilon\neq 0$, $\left|\underset{\sim}{C}'\right|\neq 0$ and:

$$\begin{aligned} \int\left|\underset{\sim}{C}'\right| \mathrm{d}x &= \frac{1}{2}\left(-c'_1+c'_4\right) \\ &= \frac{1}{2}\left(-c_1-\varepsilon+c_4+\varepsilon\right) \\ &= \frac{1}{2}\left(c_4+c_4\right) \\ &= c_4 \\ &\neq 0 . \end{aligned}$$

□

So, when $\varepsilon=0$, $\int\left|\underset{\sim}{C}'\right|=0$, but when $\varepsilon\neq 0$, $\int\left|\underset{\sim}{C}'\right|=c_4$. Now, let us suppose c_4 finite but arbitrary big and $\varepsilon\neq 0$ but infinitesimally small; thus, uncertainty on the distance is very big, even if its expected value is almost zero. To further observe this behavior, we calculated $\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right), \forall\underset{\sim}{A}=(a_1,a_2,a_3,a_4)$, $\underset{\sim}{B}=(b_1,b_2,b_3,b_4): a_i,b_i\in\{0,0.1,\ldots,1\}, i=1,\ldots,4$. In Figure 2 we can see how entropy evolves when expected distance goes to zero, showing in a black line the average entropy, in blue its the first and third quartiles, and in red the minimum and maximum entropy[1]. Except for the case of the minimum , the remaining curves show what we have already proved in Proposition 3, i.e., entropy does not go to zero when expected distance does so; but not only this, it grows, and only becomes zero for the particular case of $P\left(\left|\underset{\sim}{A},\underset{\sim}{B}\right|\right)=0$.

4 MANHATTAN DISTANCE WITH POSITIVE CORRECTION

The main goal is, therefore, to have a new definition of the absolute value that takes the positive aspects

[1]Irregularities in Figure 2 come from the discrete way in which fuzzy numbers were generated. This behavior should disappear by increasing granularity, but in turn, this increases computational complexity.

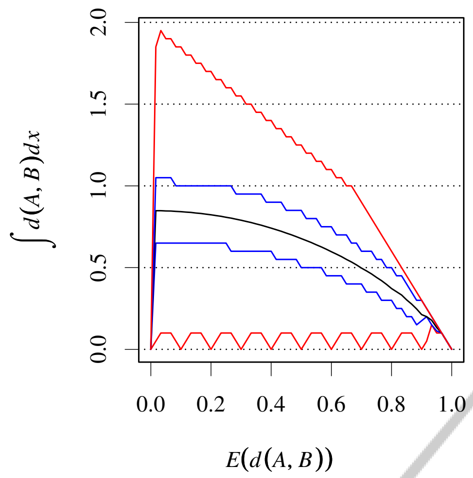


Figure 2: Entropy of the Manhattan distance as a function of its GMIR, in the interval $[0,1]$.

of both, Definitions 13 and 15, i.e., that the absolute value of a TrFN removes negative values, keeping the same expected value of the TrFN obtained through (8). For this, we will define what we have termed as "positive correction".

Definition 16. *The "positive correction" $\overset{\mapsto}{\underset{\sim}{A}}$ of a TrFN $\underset{\sim}{A} = (a_1, a_2, a_3, a_4)$ is defined as:*

$$\overset{\mapsto}{\underset{\sim}{A}} = \begin{cases} (a_1,a_2,a_3,a_4), & \text{if } a_1 \geq 0, \\ (0,a_2-\gamma_1,a_3-\gamma_1,a_4-\gamma_1), & \text{if } a_1<0 \text{ and } \frac{a_1}{5}+a_2 \geq 0, \\ (0,0,a_3-\gamma_2,a_4-\gamma_2), & \text{if } a_2+\frac{a_1}{5}<0 \\ & \text{and } \frac{a_1}{3}+\frac{2a_2}{3}+a_3 \geq 0, \\ (0,0,0,a_4-\gamma_3), & \text{if } \frac{a_1}{3}+\frac{2a_2}{3}+a_3<0, \\ (0,0,0,0), & \text{else.} \end{cases} \tag{21}$$

Proposition 4. *For a TrFN $\underset{\sim}{A} = (a_1, a_2, a_3, a_4)$, such that $a_1 < 0$ and $a_2 + \frac{a_1}{5} \geq 0$, then:*

$$\gamma_1 = -\frac{a_1}{5}.$$

Proof. Let us recall the two conditions that the positive correction of a TrFN should meet: (i) there should be no negative values, i.e., $\overset{\mapsto}{a_1} = 0$, and (ii) the expected value according to (8) must be kept the same. To comply with both conditions we must establish the following equality:

$$\begin{aligned} E\left(\underset{\sim}{A}\right) &= E\left(\overset{\mapsto}{\underset{\sim}{A}}\right) \\ \frac{a_1+2a_2+2a_3+a_4}{6} &= \frac{2(a_2-\gamma_1)+2(a_3-\gamma_1)+(a_4-\gamma_1)}{6} \\ \frac{a_1}{6} &= -\frac{5\cdot\gamma_1}{6} \\ \gamma_1 &= -\frac{a_1}{5}. \end{aligned}$$

Due to the first condition and by definition of TrFN $a_2 - \gamma_1 \geq 0$, then $a_2 + \frac{a_1}{5} \geq 0$. $\square$

Proposition 5. *For a TrFN $\underset{\sim}{A} = (a_1, a_2, a_3, a_4)$, such that $a_2 + \frac{a_1}{5} < 0$ and $\frac{a_1}{3} + \frac{2\cdot a_2}{3} + a_3 \geq 0$, then:*

$$\gamma_2 = -\frac{a_1}{3} - \frac{2\cdot a_2}{3}.$$

Proof. Again, we state the equality:

$$E\left(\underset{\sim}{A}\right) = E\left(\overset{\mapsto}{\underset{\sim}{A}}\right),$$

but this time $\overset{\mapsto}{a_1} = \overset{\mapsto}{a_2} = 0$ for $\overset{\mapsto}{\underset{\sim}{A}}$, thus:

$$\begin{aligned} \frac{a_1+2a_2+2a_3+a_4}{6} &= \frac{2(a_3-\gamma_2)+(a_4-\gamma_2)}{6} \\ \frac{a_1+2a_2}{6} &= -\frac{\gamma_2}{2} \\ \gamma_2 &= -\frac{a_1}{3} - \frac{2a_2}{3}. \end{aligned}$$

Due to the first condition of the positive correction and by definition of TrFN $a_3 - \gamma_2 \geq 0$, then $\frac{a_1}{3} + \frac{2a_2}{3} + a_3 \geq 0$. $\square$

Proposition 6. *For a TrFN $\underset{\sim}{A} = (a_1, a_2, a_3, a_4)$, such that $\frac{a_1}{3} + \frac{2a_2}{3} + a_3 < 0$ then:*

$$\gamma_3 = -x_1 - 2x_2 - 2x_3.$$

Proof. One last time, we state the equality:

$$E\left(\underset{\sim}{A}\right) = E\left(\overset{\mapsto}{\underset{\sim}{A}}\right),$$

only this time $\overset{\mapsto}{a_1} = \overset{\mapsto}{a_2} = \overset{\mapsto}{a_3} = 0$ for $\overset{\mapsto}{\underset{\sim}{A}}$, thus:

$$\begin{aligned} \frac{a_1+2a_2+2a_3+a_4}{6} &= \frac{(a_4-\gamma_3)}{6} \\ \frac{a_1+2a_2+2a_3}{6} &= -\frac{\gamma_3}{6} \\ \gamma_3 &= -a_1 - 2a_2 - 2a_3. \end{aligned}$$

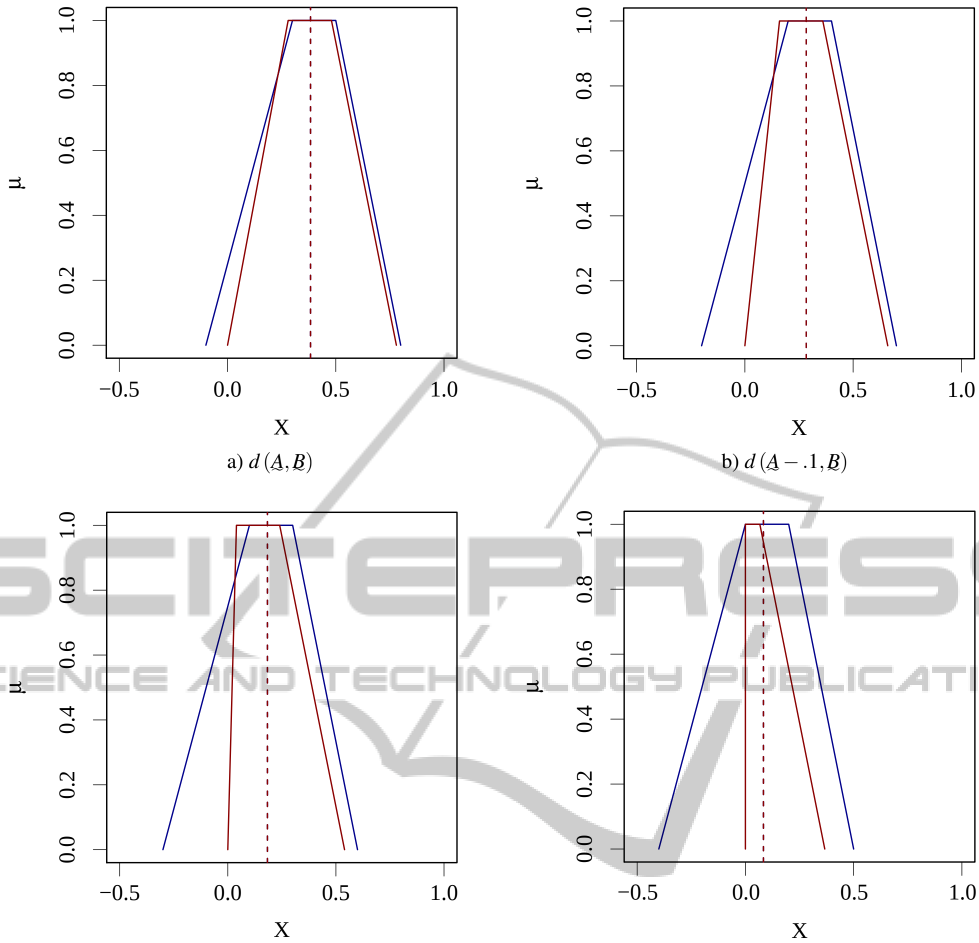
Due to the first condition of the positive correction and by definition of TrFN $a_4 - \gamma_3 \geq 0$, then $a_1 + 2a_2 + 2a_3 + a_4 \geq 0$. $\square$

Once defined the positive correction of a TrFN, we present the new absolute value of a TrFN as well as the Manhattan distance for two TrFNs.

Definition 17. *The absolute value of a TrFN $\underset{\sim}{A}$ is defined as:*

$$\left|\underset{\sim}{A}\right| = \begin{cases} \overset{\mapsto}{\underset{\sim}{A}}, & \text{if } E\left(\underset{\sim}{A}\right) > 0, \\ 0, & \text{if } E\left(\underset{\sim}{A}\right) = 0, \\ \overset{\mapsto}{\underset{\sim}{A}^-}, & \text{if } E\left(\underset{\sim}{A}\right) < 0. \end{cases} \tag{22}$$

Figure 3 shows that by using the positive correction in the absolute value, the uncertainty of the Manhattan distance goes to a maximum, top red line, when the GMIR is close to 0.5.



a) $d(\underset{\sim}{A},\underset{\sim}{B})$ b) $d(\underset{\sim}{A}-.1,\underset{\sim}{B})$

c) $d(\underset{\sim}{A}-.2,\underset{\sim}{B})$ d) $d(\underset{\sim}{A}-.3,\underset{\sim}{B})$

Figure 4: Manhattan distance calculated with the absolute value of Chen and Wang (in blue) and with the positive correction (in red), as well as the expected value calculated through the GMIR (dashed red line).

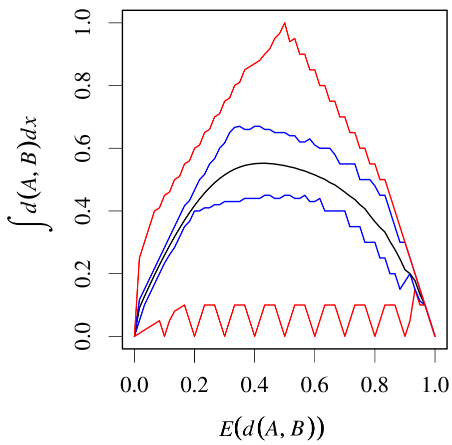


Figure 3: Entropy of the Manhattan distance with positive correction as a function of its GMIR, in the interval $[0,1]$.

Proposition 7. *For any two TrFN* $\underset{\sim}{A} = (a_1,a_2,a_3,a_4), \underset{\sim}{B} = (b_1,b_2,b_3,b_4)$ *such that* $a_1 \geq 0$, $b_1 \geq 0$, $a_4 \leq 1$, $b_4 \leq 1$, *and* $E(\underset{\sim}{A}) \geq E(\underset{\sim}{B})$,

$$\underset{d_H(\underset{\sim}{A},\underset{\sim}{B})}{\arg\max} \int_0^1 d_H(\underset{\sim}{A},\underset{\sim}{B})\, dx = (0,0,1,1)\,.$$

Proof. Given that for any two TrFN $\underset{\sim}{A}$ and $\underset{\sim}{B}$, as defined by the proposition, by Definitions 16 and 17 $d_H(\underset{\sim}{A},\underset{\sim}{B})$ is always a TrFN such that $\underset{\sim}{C} = d_H(\underset{\sim}{A},\underset{\sim}{B}) = (c_1,c_2,c_3,c_4)$ with $c_1 \geq 0$ and $c_4 \leq 1$, then:

$$\max\left(\int_0^1 \underset{\sim}{C}(x)\, \mathrm{d}x\right) = 1\,.$$

For a TrFN, this integral yields:

$$\begin{aligned}\int_0^1 \underset{\sim}{C}\, \mathrm{d}x &= \frac{c_2-c_1}{2} + c_3 - c_2 + \frac{c_4-c_3}{2} \\ &= \frac{-c_1-c_2+c_3+c_4}{2}\,. \end{aligned} \quad (23)$$

It is straightforward to see that (23) maximizes for $\underset{\sim}{C} = (0,0,1,1)$, and that $E(\underset{\sim}{C}) = 0.5$. □

We can also see in Figure 3, that the main objective pursued, i.e., that uncertainty goes to zero with the expected value of the distance, is achieved with the positive correction of the Manhattan distance. In Figure 4 we can see an example calculating $\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)$ for $\underset{\sim}{A} = (.3, .6, .7, .9)$ and $\underset{\sim}{B} = (.1, .2, .3, .4)$, using Chen and Wang's absolute value (in blue) and the positive correction (in red). We can see how the positive correction removes the negatives values of $\underset{\sim}{d_H}\left(\underset{\sim}{A},\underset{\sim}{B}\right)$, but maintains its GMIR (dashed red line), accomplishing the second objective originally proposed.

5 CONCLUSIONS

In this paper, we have discussed on the definition of the absolute value of a TrFN and its role on the calculation of the Manhattan distance for fuzzy vectors. The first definition studied, the one made by (Dubois and Prade, 1979), has two problems. Firstly, the resulting distance might not be a TrFN, which complicates following calculations. Secondly, depending on the the TrFN, it might overestimate the absolute value of its GMIR or expected value.

The second definition evaluated, that of (Chen and Wang, 2008), solves these problems but introduces some problems of its own. Firstly, the absolute value of a TrFN might have has negative values, which goes against all sense when modeling distances and their uncertainty. Secondly, it violates another commonsense condition of distances: its uncertainty must go to zero when the distance does too.

To solve these new problems we present a method called "positive correction", in which we remove negative values while keeping the same expected value, leaving the distance as a TrFN. The result is a distance that captures uncertainty, but that also stays close to the conception that distances have in the real world.

REFERENCES

Chen, S.-H. and Hsieh, C.-H. (1999). Graded mean integration representation of generalized fuzzy number. *Journal of Chinese Fuzzy System*, 5(2):1–7.

Chen, S.-H. and Wang, C.-C. (2008). Fuzzy distance of trapezoidal fuzzy numbers and application. *International Journal of Innovative Computing, Information and Control*, 4(6):2008.

Dubois, D. and Prade, H. (1979). Fuzzy real algebra: some results. *Fuzzy Sets and Systems*, 2:327–348.

Kaufmann, A. and Gupta, M. M. (1985). *Introduction to Fuzzy Arithmetic*. Van Nostrand Reinhold, New York.

Li, X. and Liu, B. (2008). On distance between fuzzy variables. *Journal of Intelligent and Fuzzy Systems*, 19:197–204.

Tran, L. and Duckstein, L. (2002). Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems*, 130:331–341.

Voxman, W. (1998). Some remarks on distances between fuzzy numbers. *Fuzzy Sets and Systems*, 100:353–365.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zadeh, L. (1975). The concept of a linguistic variable and its application to approximate reasoning - i. *Information Sciences*, 8:199–249.

Zimmermann, H.-J. (2005). *Fuzzy Sets: Theory and its Applications*. Springer, 4 edition.