The final version of this work can be found at doi:10.1016/j.fss.2014.10.003.

SISO Fuzzy Relational Inference Systems based on Fuzzy Implications are Universal Approximators

Sayantan Mandal and Balasubramaniam Jayaram,
Department of Mathematics,
Indian Institute of Technology Hyderabad,
Andhra Pradesh - 502 205, India
Email: {ma10p002, jbala}@iith.ac.in

Abstract

In this work, we show that single input single output (SISO) fuzzy inference systems based on Fuzzy Relational Inference (FRI) with implicative interpretation of the rule base are universal approximators under suitable choice of operations for the other components of the fuzzy system. The presented proofs make no assumption on the form or representations of the considered fuzzy implications and hence show that a much larger class of fuzzy implications other than what is typically considered in the literature can be employed meaningfully in FRIs based on implicative models. A concept of Weak Coherence is proposed, which plays an important role in enlarging the class of fuzzy implications that can be considered.

Keywords: Fuzzy Relational Inference, Fuzzy Implications, Universal Approximation

1. Introduction

The term *approximate reasoning* refers to methods and methodologies that enable reasoning with imprecise inputs to obtain meaningful outputs [15]. Fuzzy Inference Systems (FIS) form one particular type of approximate reasoning scheme involving fuzzy sets and are one of the best known applications of fuzzy logic in the wider sense. FIS have many degrees of freedom, namely, the underlying fuzzy partition of the input and output spaces, the fuzzy logic operations employed, the fuzzification and defuzzification procedures used, etc. This freedom gives rise to a variety of FIS with differing capabilities. The best known types of FIS are the Fuzzy Relational Inference (FRI) systems [28, 45], Similarity Based Reasoning (SBR) systems [24, 34, 13, 20] and the Takagi-Sugeno fuzzy systems [31, 32]. In this work, we focus only on the FRI systems.

1.1. Motivation for this work

One of the important factors considered while employing an FIS is its approximation capability. While many studies have appeared on this topic, most of them deal with FRIs where the rules are interpreted in a non-conditional way or as just aggregation of possibile configurations of the data (see Section 3 for details). When an implicative or a conditional interpretation of the rules is considered, there are only a few works that deal with their approximation properties.

While one of the earliest studies on this topic was that of Castro [10], [11], it was later on shown by Li *et al.* [23], [22] that some of the operations considered by Castro led to vacuous outputs. Li *et al.* further went on to present their own results on it. However, the scope of their work is restricted to the following three families of fuzzy implications, namely, R-implications from left-continuous t-norms, (S, N)-

and QL-implications. Recently, Štěpnička *et al.* [39] have discussed the same in a slightly more general setting. Once again, the assumptions they make on some components of the FRIs are not desirable - for instance, the requirements on the input partition make it non-Ruspini. Thus there is a need for constructing an implicative FRI based FIS that is both a universal approximator and whose components have some desirable properties.

Further, so far, all the studies have largely considered fuzzy implications obtained as the residuals of left-continuous t-norms, while Li *et al.* [23] had considered also the families of (S, N)- and QL-implications.Thus a study of approximation capabilities of FRIs based on fuzzy implications that come from a more general class of fuzzy implications is important.

1.2. Main contributions of the work

In this work, we provide a constructive proof of the universal approximation property of FRIs when an implicative model of the rule base is employed, i.e., where the operation between the antecedents and consequents is taken as a fuzzy implication. The proof makes no assumption on the form or representation of the considered fuzzy implications and hence is applicable for a much larger class of fuzzy implications other than what is typically considered in the literature. Especially, the presented proofs are applicable also to contexts where implications that come from a non-residuated lattice setting are employed. Further we have shown that the approximator function is continuous which is an important issue as discussed in [37].

1.3. Outline of the work

After presenting some important definitions from both fuzzy set theory and fuzzy logic connectives in Section 2, we describe the two main types of fuzzy rule bases used in fuzzy systems in Section 3. Following this, Section 4 is devoted to a full description of fuzzy relational inference (FRI) mechanisms. In Section 5 we present a short survey on the works and results related to universal approximation of fuzzy relational inference systems. Further we clearly specify the main contributions and scope of our work.

Relaxing the often insisted coherence of an implicative model suitably to the context of function approximation, Section 6 investigates the class of fuzzy implications that can be used in FRIs to ensure this form of weak coherence. This section also presents some well-known families of fuzzy implications that belong to the above admissible class of fuzzy implications.

Finally, Sections 7 and 8 contain the major contributions of this work, which show that FRIs employing a rather large class of fuzzy implications - which include the R-implications - are universal approximators. Section 9 presents some examples that illustrate the investigations and analysis of the previous sections. In Section 10 some concluding remarks are given.

2. Preliminaries

We assume that the reader is familiar with the classical results concerning fuzzy set theory and basic fuzzy logic connectives, but to make this work more self-contained, we introduce some notations, concepts and results employed in the rest of the work.

In this work we only consider $X \subseteq \mathbb{R}$ to be a closed and bounded interval and hence X is totally ordered, linear and compact w.r.to the usual topology on $\mathbb{R}$. However, many of the concepts below are applicable to more general sets and hence the definition is given accordingly.

2.1. Fuzzy Sets

Definition 2.1. *If X is a non-empty set then $\mathcal{F}(X)$ is the fuzzy power set of X, i.e., $\mathcal{F}(X) = \{A|A : X \to [0,1]\}$.*

Definition 2.2. *A fuzzy set A is said to be*

- normal *if there exists an $x \in X$ such that $A(x) = 1$,*

- convex *if X is a compact subset of a linear space and for any $\lambda \in [0,1]$, $x, y \in X$, $A(\lambda x + (1-\lambda)y) \geq \min\{A(x), A(y)\}$.*

Definition 2.3. *For an $A \in \mathcal{F}(X)$, the* Support, Height, Kernel *and* Ceiling *of A are denoted, respectively, as Supp A, Hgt A, Ker A and Ceil A and are defined as:*

$$\begin{aligned} \mathit{Supp}\ A &= \{x \in X | A(x) > 0\}\ , \\ \mathit{Hgt}\ A &= \sup\{A(x) | x \in X\}\ , \\ \mathit{Ker}\ A &= \{x \in X | A(x) = 1\}\ , \\ \mathit{Ceil}\ A &= \{x \in X | A(x) = \mathit{Hgt}\ A\}\ . \end{aligned}$$

A is said to be bounded *if Supp A is a bounded set. Note that for a normal fuzzy set Ker A = Ceil A and Hgt $A = 1$.*

Definition 2.4. *Let $\mathcal{P}$ be a finite collection of fuzzy sets of X, i.e, $\mathcal{P} = \{A_k\}_{k=1}^n \subseteq \mathcal{F}(X)$. $\mathcal{P}$ is said to form a* fuzzy partition *on X if*

$$X \subseteq \bigcup_{k=1}^{n} \mathit{Supp}\ A_k\ .$$

In the literature, a partition $\mathcal{P}$ of X as defined above is also called a **complete** partition. Note that there are several other approaches to and definitions of a fuzzy partition, see for instance, [9, 14, 18, 26, 27]

Definition 2.5. *A fuzzy partition $\mathcal{P} = \{A_k\}_{k=1}^n \subseteq \mathcal{F}(X)$ is said to be*

- ***consistent*** *if whenever for some k, $A_k(x) = 1$ then $A_j(x) = 0$ for $j \neq k$,*
- *a **Ruspini Partition** if*

$$\sum_{k=1}^{n} A_k(x) = 1 \ \ \textit{for every}\ \ x \in\ X. \tag{1}$$

2.2. Defuzzification

Often there is a need to convert a fuzzy set to a crisp value, a process which is called *Defuzzification.* This process of defuzzification can be seen as a mapping $d : \mathcal{F}(X) \longrightarrow X$. There are many types of defuzzification techniques available in the literature, see [30] for a good overview. The defuzzifier given in Example 2.6 will be used extensively in the sequel.

Example 2.6. *For a fuzzy set $A \in \mathcal{F}(X)$, with bounded Ceil A, the* Mean of Maxima *(*MOM*) defuzzifier gives as output the mean of all those values in X with the highest membership value, which can be mathematically expressed as*

$$\mathrm{MOM}(A) = \frac{\displaystyle\int_{\mathit{Ceil}\ A} x dx}{\displaystyle\int_{\mathit{Ceil}\ A} 1 dx}\ ,\ \ \textit{if}\ \ \int_{\mathit{Ceil}\ A} 1 dx \neq 0. \tag{2}$$

2.3. Fuzzy Logic Connectives

Note that in this work, we use the term *decreasing* and *increasing* in a non-strict sense. In other words, we call a function $t_1 : \mathbb{R} \to \mathbb{R}$ decreasing or non-increasing if $t_1(x) \geq t_1(y)$ whenever $x \leq y$. Similarly, we call a function $t_2 : \mathbb{R} \to \mathbb{R}$ increasing or non-decreasing if $t_2(x) \leq t_2(y)$ whenever $x \leq y$.

Definition 2.7 ([21]). *A binary operation $T \colon [0,1]^2 \to [0,1]$ is called a* t-norm*, if it is increasing in both variables, commutative, associative and has 1 as the neutral element.*

Definition 2.8 ([21]). *A t-norm T is called positive if $T(x,y)=0$ then either $x=0$ or $y=0$.*

Definition 2.9 ([2], Definition 1.1.1). *A function $I\colon [0,1]^2 \to [0,1]$ is called a* fuzzy implication *if it satisfies, for all $x, x_1, x_2, y, y_1, y_2 \in [0,1]$, the following conditions:*

$$\text{if } x_1 \le x_2, \text{ then } I(x_1,y) \ge I(x_2,y)\ , \text{ i.e., } I(\,\cdot\,,y) \text{ is decreasing}\ ,$$
$$\text{if } y_1 \le y_2, \text{ then } I(x,y_1) \le I(x,y_2)\ , \text{ i.e., } I(x,\,\cdot\,) \text{ is increasing}\ ,$$
$$I(0,0)=1\ , I(1,1)=1\ , I(1,0)=0\ .$$

The set of all fuzzy implications will be denoted by $\mathbb{I}$.

Definition 2.10 ([2]). *A fuzzy implication $I\colon [0,1]^2 \to [0,1]$ is said to*

- *satisfy the* left neutrality property*, if*

$$I(1,y)=y\ , \qquad y\in[0,1]\ , \tag{NP}$$

- *satisfy the* ordering property*, if*

$$I(x,y)=1 \Longleftrightarrow x\le y\ , \qquad x,y\in[0,1]\ . \tag{OP}$$

- *be a* positive fuzzy implication *if $I(x,y)>0$, for all $x,y\in(0,1]$.*

For examples of such fuzzy implications, please refer to Table 1.

Definition 2.11 ([2]). *A function $N:[0,1]\longrightarrow[0,1]$ is called a* fuzzy negation *if $N(0)=1, N(1)=0$ and N is decreasing.*

Example 2.12. *One such fuzzy negation is the Gödel negation*

$$N_{\mathbf{D1}}(x)=\begin{cases}1, & \text{if } x=0\ ,\\ 0, & \text{if } x>0\ ,\end{cases} \qquad x\in[0,1]\ . \tag{3}$$

Definition 2.13 ([2]). *Let $I\in\mathbb{I}$ be any fuzzy implication. The function $N_I:[0,1]\longrightarrow[0,1]$ defined by $N_I(x)=I(x,0)$ is a fuzzy negation and is called the* natural negation *of I.*

We denote the class of fuzzy implications that satisfy (OP) by $\mathbb{I}_{\mathbf{OP}} \subsetneq \mathbb{I}$ and the class of fuzzy implications that are positive and whose natural negation $N_I = N_{\mathbf{D1}}$ by $\mathbb{I}^{+}_{N_{\mathbf{D1}}} \subsetneq \mathbb{I}$.

3. Fuzzy IF-THEN Rule Base

Given two non-empty crisp sets $X, Y \subseteq \mathbb{R}$, a Single-Input Single-Output (SISO) fuzzy IF-THEN rulebase consists of rules of the form:

$$\textbf{IF } \tilde{x} \text{ is } A_i \textbf{ THEN } \tilde{y} \text{ is } B_i\ , \tag{4}$$

where $\tilde{x}$,$\tilde{y}$ are the linguistic variables and $A_i, B_i, i=1,2,\ldots n$ are the linguistic values taken by the linguistic variables. These linguistic values are represented by fuzzy sets in their corresponding domains, i.e., $A_i \in \mathcal{F}(X), B_i \in \mathcal{F}(Y)$.

As an example,

IF $Temperature$ is $High$ **THEN** $Fanspeed$ is $Medium$.

Here $Temperature$ and $Fanspeed$ are the linguistic variables and $High$, $Medium$ are the linguistic values taken by the linguistic variables in a suitable domain.

A fuzzy rule base (4) can be viewed in two different ways, as explained in [16],[17]. When each of the rules is viewed as a *constraint*, i.e., when the rules in (4) are combined together as

$$\begin{aligned}
&\textbf{IF } \tilde{x} \text{ is } A_1 \textbf{ THEN } \tilde{y} \text{ is } B_1,\\
&\qquad\qquad \dots\\
&\qquad\qquad \textbf{AND}\\
&\qquad\qquad \dots\\
&\textbf{IF } \tilde{x} \text{ is } A_n \textbf{ THEN } \tilde{y} \text{ is } B_n,
\end{aligned} \tag{5}$$

we have the conditional form (IF-THEN) of the rules. On the other hand, each of the rules can also be viewed as just pieces of data giving possible configurations or positive information, in which case they are combined as follows:

$$\begin{aligned}
&\tilde{x} \text{ is } A_1 \textbf{ AND } \tilde{y} \text{ is } B_1,\\
&\qquad \dots\\
&\qquad \textbf{OR}\\
&\qquad \dots\\
&\tilde{x} \text{ is } A_n \textbf{ AND } \tilde{y} \text{ is } B_n.
\end{aligned} \tag{6}$$

3.1. Fuzzy Relations that model the Rule Base (5) *and* (6)

In fuzzy relational inference mechanisms (see Section 4 below), fuzzy relations $R : X \times Y \to [0,1]$ are employed to represent the rule base (5) and (6). Two of the commonly employed fuzzy relations are the following: For any $x \in X, y \in Y$,

$$\hat{R}_{\to}(x,y) = \bigwedge_{i=1}^{n} (A_i(x) \longrightarrow B_i(y)) \ , \tag{7}$$

$$\check{R}_{\star}(x,y) = \bigvee_{i=1}^{n} (A_i(x) \star B_i(y)) \ , \tag{8}$$

where $\longrightarrow$ is taken as a fuzzy implication and $\star$ as a t-norm.

Note that the fuzzy relation $\hat{R}_{\to}$ captures the conditional form (5) of the given rules, while the relation $\check{R}_{\star}$ captures the Cartesian product form (6) of the rules. For more on the semantics of $\check{R}_{\star}$ and $\hat{R}_{\to}$ we refer the readers to [17].

4. Fuzzy Relational Inference mechanism

Given a rule base of the form (5) or (6) and an input " $\tilde{x}$ is A' ", the main objective of a fuzzy inference mechanism is to find a *meaningful* B' such that " $\tilde{y}$ is B' ". While many types of fuzzy inference mechanisms have been proposed in the literature we restrict this study only to fuzzy relation based inference mechanisms.

The inference mechanism in a fuzzy relational inference (FRI) can be expressed as follows:

$$B' = f_R^{@}(A') = A'@R, \tag{FRI-R}$$

where $A' \in \mathcal{F}(X)$ is the input, the relation $R \in \mathcal{F}(X \times Y)$ represents or models the rule base, $B' \in \mathcal{F}(Y)$ is the obtained output and @ is called the *composition operator*, which is a mapping $@ \colon \mathcal{F}(X) \times \mathcal{F}(X \times Y) \to \mathcal{F}(Y)$.

4.1. Two main types of FRIs

One of the two main FRIs is the Compositional Rule of Inference (CRI) proposed by Zadeh [44] and given as:

$$B'(y) = f_R^{\circ}(A')(y) = \bigvee_{x \in X} [A'(x) \star R(x,y)], \qquad y \in Y, \tag{CRI-R}$$

where $\star$ is a t-norm. The operator $\circ$ is also known as the sup $-T$ composition where T is a t-norm. Later Pedrycz [28] proposed another FRI mechanism based on the Bandler-Kohout Subproduct composition given as:

$$B'(y) = f_R^{\triangleleft}(A')(y) = \bigwedge_{x \in X} [A'(x) \longrightarrow R(x,y)], \qquad y \in Y, \tag{BKS-R}$$

with $\longrightarrow$ interpreted as a fuzzy implication. The operator $\triangleleft$ is also known as the inf $-I$ composition where I is a fuzzy implication. Note that $f_R^{\circ}(A')$ is also known as the *direct image* of A' over R, while $f_R^{\triangleleft}(A')$ is called the *sub-direct image* of A' over R [39].

4.2. FRI with Singleton input

Since we deal with universal approximation capability, the inputs are crisp and hence we need to suitably fuzzify the crisp input to a fuzzy input.

If $x_0 \in X$ is a crisp input, then it is suitably *fuzzified*, i.e., a fuzzy set $A' \in \mathcal{F}(X)$ is suitably constructed from x_0. Commonly, the following *singleton* fuzzifier is employed:

$$A'(x) = \begin{cases} 1, & x = x_0 \ , \\ 0, & x \neq x_0 \ . \end{cases}$$

With the above input A', the FRI mechanism (FRI-R) reduces to

$$B'(y) = R(x_0, y), \qquad y \in Y, \tag{FRI-R-Singleton}$$

for any t-norm $\star$ in case of (CRI-R) and any implication I satisfying (NP) in case of (BKS-R). Thus in the case of a singleton input both the (CRI-R) and (BKS-R) are essentially the same (provided $\longrightarrow$ in (BKS-R) satisfies (NP)) and the output is fully dependent on the model of the rule base R. In other words, $f_R^{\circ} \equiv f_R^{\triangleleft}$ and hence the composition $\circ$ or $\triangleleft$ - when the I in $\triangleleft =$ inf $-I$ composition satisfies (NP) - does not play any role.

4.3. Scope of this work

We denote an FRI with singleton input as a quadruple $\mathbb{F} = \left(\mathcal{P}_X, \mathcal{P}_Y, R, d\right)$, where $\mathcal{P}_X = \{A_i\}$ and $\mathcal{P}_Y = \{B_i\}$ correspond to the input and output fuzzy partitions on X and Y, respectively, R is the fuzzy relation modeling the rule base and d is the defuzzifier used to obtain a crisp output from the obtained B' in (FRI-R-Singleton). Thus given an $\mathbb{F}$ the overall inference can be seen as a function $g : X \to Y$ as follows:

$$g(x') = d(B'(\cdot)) = d(R(x', \cdot)) \ , \ x' \in X. \tag{9}$$

In the literature, g is also known as the system function of a given $\mathbb{F}$, see for instance, [22, 23].

In this paper we deal only with the implicative form of the rule base, i.e., the antecedents of the rules are related to their consequents using a fuzzy implication and hence fix $R = \hat{R}_{\rightarrow}$ in the sequel.

Further, we consider the following generalised form of $\hat{R}_{\rightarrow}$, where T is any t-norm, not necessarily the minimum t-norm:

$$R_{\rightarrow}^{T}(x,y) = T_{i=1}^{n}(A_i(x) \longrightarrow B_i(y)), \tag{Imp-$R_{\rightarrow}^{T}$}$$

Thus this work deals with FRIs of the form $\mathbb{F}_{\rightarrow}^{T} = \left(\mathcal{P}_X, \mathcal{P}_Y, R_{\rightarrow}^{T}, d\right) \subseteq \mathbb{F}$.

5. FRIs as Universal Approximators

5.1. Universal Approximation results with FRI

In this subsection, we only briefly recall some of the important works dealing with the approximation properties of FRIs and refer the readers to the excellent review of Tikk *et al.* [33] for more details and the other recent works, for instance [29], [43] and the references therein.

The earliest works to appear on this topic dealt with FRIs where $R = \check{R}_{\star}$ and hence can be considered to have assumed a Cartesian product interpretation of the fuzzy rules, see Wang [41] and Zeng & Singh [46].

It was Castro [10] who was the first to deal with the approximation properties of FRIs that employed $\hat{R}_{\rightarrow}$. However, as was already pointed out by Li *et al.* *Remark 2.4*, [23], Castro has considered an FRI as given below:

$$B'(y) = \bigvee_{j} (B'_j(y)) = \bigvee_{j} (A_j(x_0) \longrightarrow B_j(y)) \ ,$$

which is clearly not an appropriate model to work with, since under most practical settings for any given $x_0 \in X$ there will always exist a rule with an antecedent A_{i_0} such that $A_{i_0}(x_0) = 0$ and since $0 \longrightarrow b = 1$ for any $b \in [0, 1]$, when the maximum t-conorm is used to aggregate the individual outputs one always obtains that $B'(y) = 1$ for all $y \in Y$. Note that this is the case when the input partitions are of the Ruspini type - a property that is normally both practical and desirable.

In the same work, after pointing out the above, Li *et al.* (see *Theorem 3.4*, [23]) have given a constructive proof of the approximation capability of an FRI with $R = \hat{R}_{\rightarrow}$. However, the scope of their work is restricted to the following three families of fuzzy implications, namely, R-implications from left-continuous t-norms, (S, N)- and QL-implications. Further, many of the results are without complete and correct proofs, thus making a deeper understanding of the approach difficult.

Perfilieva and Kreinovich [29] have discussed approximation capability of fuzzy systems that reflect the CNF-DNF duality. However, the considered / constructed partitions are not 'fuzzy' and hence the constructed rule base contains antecedents and consequents that are crisp sets. Further, they make an implicit assumption that the considered implications can be written as a generalization of the classical material implication, which in the context of fuzzy logic connectives is equivalent to assuming that the considered implication is an (S, N)-implication. While this assumption is valid in their context, since the relations R_{CNF} are crisp and hence only need to deal with $\{0, 1\}$ values, in general, this is not true when we consider truth-values over the entire $[0, 1]$ interval.

Recently, Štěpnička *et al.* in [39] considered an FRI with $R = R^{\otimes}_{\rightarrow_*}$ where $\otimes$ is the Łukasiewicz t-norm $T_{\mathbf{LK}}(x, y) = \max(0, x + y - 1)$ and $\rightarrow_*$ is any residuated implication obtained from a left-continuous t-norm $*$, which can be different from $T_{\mathbf{LK}}$. They have shown that the FRIs $\mathbb{F}^{\otimes}_{\rightarrow_*} = \mathbb{F}^{T_{\mathbf{LK}}}_{\rightarrow_*} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^{T_{\mathbf{LK}}}_{\rightarrow_*}, \mathrm{MOM}\right)$ are universal approximators. Their result is true for any continuous function f but the A_i's do not form a Ruspini partition which is normal and desirable in practical settings.

5.2. Main contribution of this work

In this work, we show that FRIs of the form $\mathbb{F}^{T}_{\rightarrow} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^{T}_{\rightarrow}, d\right)$ are universal approximators. Moreover it has also been shown that the system function g of the given $\mathbb{F}^{T}_{\rightarrow}$ is continuous. A concept of *weak coherence* is proposed, which plays an important role in enlarging the class of fuzzy implications that can be considered. The proof is general enough for a large class of fuzzy implications and is valid for any continuous function, not necessarily monotonic and the partitions used are of the Ruspini type. Thus, we believe that these results are very much applicable in most of the practical and desirable contexts [17, 39, 40].

6. Weak Coherence and Implicative Models

6.1. A Weaker form of Coherence

Dubois *et al.* [17] defined the concept of coherence for an implicative model $\hat{R}_{\rightarrow}$ (see (7)) of a rule base as follows, which is suitably modified to fit into our notation.

Definition 6.1 ([12], [17]). *Given a rule base* (5)*, a fuzzy relation* $R^T_{\rightarrow}(x,y)$*, as in* (Imp-$R^T_{\rightarrow}$) *modelling this rule base, is coherent if for any* $x \in X$ *there exist* $y \in Y$ *such that* $R^T_{\rightarrow}(x,y) = 1$*.*

The coherence property states that for any x the final fuzzy output B' should be normal, i.e., *Ker* $B' \neq \emptyset$. Coherence of an implicative model of a rule base is very much dictated by the semantics involved [17]. Further, it is essential when using defuzzification techniques that are dependent on the kernel to be non-empty.

However, there exist other *reasonable* defuzzification methods that do not depend on the kernel of the output fuzzy set and, further, in the setting of function approximation, as is the case here, perhaps there is an arguable justification to not to insist on this otherwise extremely important property. Relaxing this property we define the following weaker form of coherence.

Definition 6.2. *For a given rule base* (5)*, a fuzzy relation* $R^T_{\rightarrow}(x,y)$ *is said to be* weakly coherent *if for any* $x \in X$ *there exist* $y \in Y$ *such that* $R^T_{\rightarrow}(x,y) > 0$*.*

From (FRI-R-Singleton) and (Imp-$R^T_{\rightarrow}$), we have the following:

$$\begin{aligned} B'(y) &= R^T_{\rightarrow}(x_0,y) = T^n_{i=1}(A_i(x_0) \longrightarrow B_i(y)) \\ &= T\big(A_1(x_0) \longrightarrow B_1(y), A_2(x_0) \longrightarrow B_2(y), \ldots, A_n(x_0) \longrightarrow B_n(y)\big). \end{aligned}$$

Now if the antecedent fuzzy sets are normal and form a Ruspini partition, then x_0 intersects atmost two fuzzy sets say, A_m, A_{m+1}. Then the above reduces to

$$B'(y) = T(A_m(x_0) \longrightarrow B_m(y), A_{m+1}(x_0) \longrightarrow B_{m+1}(y)) = T(B'_m(y), B'_{m+1}(y)) \ , \tag{10}$$

where B'_m and B'_{m+1} are the fuzzy sets B_m and B_{m+1} modified by the fuzzy implication $\longrightarrow$ with $A_m(x_0)$ and $A_{m+1}(x_0)$, respectively.

It is clear that for B' to be non-empty the supports of B'_m and B'_{m+1} should intersect, i.e., a necessary condition is

$$\text{Supp } B'_m \bigcap \text{Supp } B'_{m+1} \neq \emptyset \ . \tag{11}$$

Further, the choice of the t-norm T in (10) should also be made accordingly. For instance, a sufficient condition on T is that it should be positive (see Definition 2.8). Then from (11) we see that B' is non-empty.

While coherence insists that the kernels of B'_m and B'_{m+1} should intersect, the weak coherence defined above relaxes this to a mere intersection of their supports. It should be noted that while relaxing coherence to weak coherence does expand the set of fuzzy implications that can be considered in $\hat{R}_{\rightarrow}$, it still does not encompass the whole set of fuzzy implications $\mathbb{I}$.

In the following, we discuss the class of fuzzy implications that can be considered for an FRI with $R^T_{\rightarrow}$ to be at least weakly coherent. This leads us to study the effect of using fuzzy implications to modify fuzzy sets.

6.2. Fuzzy Sets modified by Fuzzy Implications

From the above section it is clear that to ensure weak coherence, we need to deal with fuzzy sets that are modified by a fuzzy implication. Thus studying the properties of such modified fuzzy sets is important and we proceed to do this in the section.

Definition 6.3. *Let* $C \in \mathcal{F}(X)$ *and* $I \in \mathbb{I}$ *be any fuzzy implication. We say that a* $C^I_\alpha \in \mathcal{F}(X)$ *is the* modification or modified fuzzy set of C by I at a given $\alpha \in [0,1]$ *if*

$$C^I_\alpha(x) = I(\alpha, C(x)), \ x \in X. \tag{12}$$

Since in this work we consider modification only by a fuzzy implication, we often use the simpler term *modified fuzzy set* without any explicit mention of either I or the $\alpha \in [0,1]$.

The following results show that modification by an $I \in \mathbb{I}$ preserves convexity and also gives some relations between the supports of the original and modified fuzzy sets when an $I \in \mathbb{I}$ is used.

Proposition 6.4. *For a convex fuzzy set C, a fuzzy implication I and any $\alpha \in [0,1]$, $C^I_\alpha = I(\alpha, C)$ is also convex.*

Proof. C being a convex fuzzy set, for all $\lambda \in [0,1]$ and $x, y \in X$, we have $C(\lambda x + (1-\lambda)y) \geq C(x) \wedge C(y)$. Now for any $\alpha \in [0,1]$ we have the following:

$$\begin{aligned}
& C(\lambda x + (1-\lambda)y) \geq C(x) \wedge C(y) \\
\Longrightarrow & I(\alpha, C(\lambda x + (1-\lambda)y)) \geq I(\alpha, C(x) \wedge C(y)) \\
\Longrightarrow & I(\alpha, C(\lambda x + (1-\lambda)y)) \geq I(\alpha, C(x)) \wedge I(\alpha, C(y)) \\
\Longrightarrow & C^I_\alpha(\lambda x + (1-\lambda)y) \geq C^I_\alpha(x) \wedge C^I_\alpha(y).
\end{aligned}$$

This proves that the modified fuzzy set $C^I_\alpha = I(\alpha, C)$ is convex.

Remark 6.5. *In fact, the above result is true for any increasing function t. In Proposition 6.4, $t(C) = C^I_\alpha = I(\alpha, C)$, where $\alpha \in [0,1]$ is a constant.*

Proposition 6.6. *Let C be a bounded, normal, continuous convex fuzzy set, $I \in \mathbb{I}$ and $\alpha \in (0,1)$. Consider the following inclusion relating the supports of C and its modified set C^I_α:*

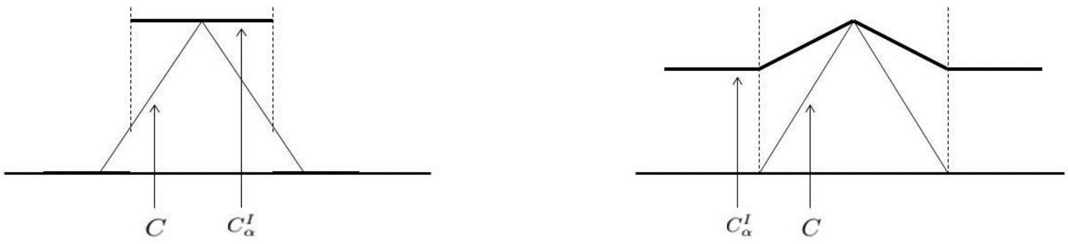
$$\textit{Supp } C^I_\alpha \supseteq \textit{Supp } C\ . \qquad (13)$$

(i) *If I is a non-positive fuzzy implication, then there exists an $\alpha \in [0,1]$ such that* (13) *is not valid.*

(ii) *For a given $I \in \mathbb{I}$, let $A_I = \{x \in [0,1] | I(x,0) = 0\}$ and let $\delta = \inf A_I$.*
   - (a) *If $\alpha < \delta$, then* (13) *is valid always.*
   - (b) *If $\alpha > \delta$, then* (13) *is valid only if I is positive.*
   - (c) *Let $\alpha = \delta$. If $\delta \in A_I$, then* (13) *is valid only if I is positive, while* (13) *holds for any $I \in \mathbb{I}$ if $\delta \notin A_I$.*

Proof. (i) Since I is non-positive, there exists some $x_0, y_0 \in (0,1)$ such that $I(x_0, y_0) = 0$. By the monotonicity of I we have that for $\alpha \in [x_0, 1]$ and $y \in [0, y_0]$, $I(\alpha, y) = 0$. Since C is continuous, normal and convex, there will exist a $U \subseteq X$ such that $C(x) \leq y_0$ on U. If we take $\alpha \in [x_0, 1]$ then $C^I_\alpha(x) = 0$ for all $x \in U$, i.e., Supp $C^I_\alpha \subsetneq$ Supp C. *For a graphical illustration where $I = I_{\mathbf{RS}}$, the Rescher implication, see Fig. 1(a).*

(ii) Let $\delta = \inf A_I = \inf\{x \in [0,1] | I(x,0) = 0\}$. Note that for any $I \in \mathbb{I}$ we have $I(1,0) = 0$ and hence $\{x \in [0,1] | I(x,0) = 0\} \neq \emptyset$. Let us consider an $\alpha \in [0,1]$.
   - (a) Let $\alpha < \delta$. Then $I(\alpha, 0) > 0$ and by the monotonicity of I, we have $I(\alpha, \beta) > 0$ for any $\beta \in [0,1]$. On the one hand, if $x \in X \setminus$ Supp C, then $C(x) = 0$ and $C^I_\alpha(x) > 0$. On the other hand, when $x \in$ Supp C, then $C(x) > 0$ and $C^I_\alpha(x) > 0$. Thus it is clear that Supp $C \subseteq$ Supp C^I_α and (13) holds. *For a graphical illustration where $I = I_{\mathbf{RC}}$, the Reichenbach implication, see Fig. 1(b).*
   - (b) Let $\alpha > \delta$. Once again, by the monotonicity of I, we have $I(\alpha, 0) = 0$. If $x \in X \setminus$ Supp C, then $C(x) = 0$ and hence $C^I_\alpha(x) = 0$. If $x \in$ Supp C then $C(x) \in (0,1]$. In fact, by the continuity and normality of C, for any $\beta \in (0,1)$ there exists an $x \in$ Supp C such that $C(x) = \beta$. Now, if I is not positive, i.e., if there exists a $\beta \in (0,1)$ such that $I(\alpha, \beta) = 0$ then for all $x \in$ Supp C such that $C(x) \leq \beta$ we have that $C^I_\alpha(x) = 0$. Thus to ensure that (13) holds we need an I which is positive.
   - (c) Let $\alpha = \delta$. If $\delta \in A_I$, then $I(\delta, 0) = I(\alpha, 0) = 0$ and hence it reduces to the case (b) above. If $\delta \notin A_I$, then $I(\delta, 0) = I(\alpha, 0) > 0$ and hence it reduces to the case (a) above.



(a) Supp $C^I_\alpha \subsetneq$ Supp C (b) Supp $C \subseteq$ Supp C^I_α

Figure 1: Inclusions between the supports of the original and modified fuzzy sets, when (a) I is non-positive, (b) I is positive and $N_I \neq N_{\mathbf{D1}}$.

Remark 6.7. *Note that, from Proposition 6.6 we see that whenever $\delta < 1$, to ensure that* (13) *holds we need an $I \in \mathbb{I}$ that is positive. If an $I \in \mathbb{I}$ which is positive and whose $N_I = N_{\mathbf{D1}}$ is used to modify C above, then the supports of C^I_α, C are equal, i.e., Supp $C^I_\alpha =$ Supp C, for all $\alpha \in (0, 1]$.* For a graphical illustration where $I = I_{\mathbf{GD}}$, the Gödel implication, see Fig. 2.

Also, note that when $I \in \mathbb{I}$ is positive but $N_I \neq N_{\mathbf{D1}}$ then the modified fuzzy set may have infinite support, in which case (13) *holds trivially (Fig. 1(b)).*

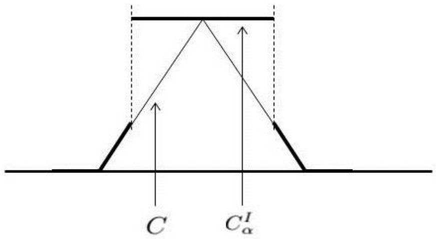


Figure 2: Supp $C \subseteq$ Supp C^I_α - When I is positive and $N_I = N_{\mathbf{D1}}$ - see Remark 6.7

6.3. Types of Fuzzy Implications Considered

From Section 6.1 above we know that for an $R^T_{\rightarrow}$ to ensure weak coherence, we need the support of the output fuzzy sets B'_m and B'_{m+1} - which are the modified fuzzy sets of B_m, B_{m+1} using a fuzzy implication $I \in \mathbb{I}$ - to intersect. Also, it can be seen from Section 6.2 that when we use a non-positive fuzzy implication the supports of these modified fuzzy sets can shrink and hence there is a possibility that the intersection of their supports is *empty*, which is not desirable. Hence to ensure weak coherence at the least, we see that the class of implications I that can be considered should be restricted.

Since in most practical settings we deal only with fuzzy sets that are bounded, continuous, convex and that which often form a Ruspini partition, it is sufficient to consider fuzzy implications $I \in \mathbb{I}$ that either

- satisfy the ordering property (OP), i.e., $I \in \mathbb{I}_{\mathbf{OP}}$, in which case often we can ensure even coherence [39], or
- are positive with $N_I = N_{\mathbf{D1}}$, i.e., $I \in \mathbb{I}^{+}_{N_{\mathbf{D1}}}$, in which case we can ensure at least a weak coherence.

Thus, in the following sections we will deal with rules modeled by fuzzy relations $R^T_{\rightarrow}$ where the fuzzy implication $\rightarrow$ either satisfies (OP) or is positive with or without (OP) but whose natural negation $N_I = N_{\mathbf{D1}}$, the Gödel negation (3).

Remark 6.8. *Note that the properties* (OP)*, positivity and* $N_I = N_{\mathbf{D1}}$ *are not mutually exclusive. Table 1 lists some fuzzy implications illustrating the same.*

Implications	(OP)	Positive	$N_I = N_{\mathbf{D1}}$
$I_{\mathbf{GG}}(x,y) = \min(1, \frac{y}{x})$	✓	✓	✓
$I_{\mathbf{LK}}(x,y) = \min(1, 1-x+y)$	✓	✓	×
$I_{\mathbf{YG}}(x,y) = \min(1, y^x)$	×	✓	✓
$I_{\mathbf{RC}}(x,y) = 1-x+xy$	×	✓	×
$I_{\mathbf{RS}}(x,y) = \begin{cases} 1, & \text{if } x \le y \\ 0, & \text{if } x > y \end{cases}$	✓	×	✓
$I(x,y) = \begin{cases} 0, & \text{if } (x,y) \in [0.7,1] \times [0,0.6] \\ 0.5, & \text{if } (x,y) \in [0.4,0.7] \times [0,0.6] \\ 1, & \text{otherwise} \end{cases}$	×	×	×
$I(x,y) = \begin{cases} 1, & \text{if } x \le y \\ 0.5, & \text{if } x > y \text{ and } x \in [0,0.5) \\ 0, & \text{if } (x,y) \in [0.5,1) \times [0,0.5) \\ 0.5, & \text{if } x > y \text{ and } y \in [0.5,1) \end{cases}$	✓	×	×
$I(x,y) = \begin{cases} 1, & \text{if } x = 0 \text{ or } y = 1 \\ 0, & \text{if } x > 0 \text{ or } y < 1 \end{cases}$	×	×	✓

Table 1: Fuzzy Implications that satisfy some or all of the properties of (OP), positivity and $N_I = N_{\mathbf{D1}}$.

6.4. Families of Fuzzy Implications that belong to $\mathbb{I}_{\mathbf{OP}} \cup \mathbb{I}^{+}_{N_{\mathbf{D1}}}$

In fact, many established families of fuzzy implications fall in either of the above two classes. For the definitions and the properties these families satisfy, please refer to the monograph [2].

- Let $\mathbb{I}_{\mathbb{T}_{\mathbb{BC}}}$ denote the set of all R-implications obtained from border continuous t-norms. Then every $I \in \mathbb{I}_{\mathbb{T}_{\mathbb{BC}}}$ satisfies (OP) ([3], Proposition 5.8). Further, the set of all R-implications obtained from left-continuous t-norms $\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}} \subsetneq \mathbb{I}_{\mathbb{T}_{\mathbb{BC}}}$ and hence we have that
$$\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}} \subsetneq \mathbb{I}_{\mathbb{T}_{\mathbb{BC}}} \subsetneq \mathbb{I}_{\mathbf{OP}} \ .$$

- Let $\mathbb{I}_{\mathbb{F}}$ denote the set of all f-implications proposed by Yager [42]. Further, let us denote by $\mathbb{I}_{\mathbb{F},\infty} \subsetneq \mathbb{I}_{\mathbb{F}}$ the set of f-implications that are generated from generators such that $f(0) = \infty$. Every $I \in \mathbb{I}_{\mathbb{F},\infty}$ is positive and their natural negation is the Gödel negation (see [19], [1], Proposition 2), i.e., $N_I = N_{\mathbf{D1}}$. Thus
$$\mathbb{I}_{\mathbb{F},\infty} \subsetneq \mathbb{I}^{+}_{N_{\mathbf{D1}}}.$$

- If $\mathbb{I}_{\mathbb{G}}$ denotes the set of all g-implications, proposed by Yager [42], then every $I \in \mathbb{I}_{\mathbb{G}}$ is positive and $N_I = N_{\mathbf{D1}}$ (see [1], Proposition 4). Thus
$$\mathbb{I}_{\mathbb{G}} \subsetneq \mathbb{I}^{+}_{N_{\mathbf{D1}}}.$$

- If $\mathbb{I}^{*}_{\mathbb{S}}$ denote the set of all $(S,N)-$ implications such that $N = N_S$, the natural negation of S, is a strong negation and the pair (S, N_S) is such that $S(N_S(x), x) = 1, \ \ x \in [0,1]$ then every $I \in \mathbb{I}^{*}_{\mathbb{S}}$ satisfies (OP) ([3], Theorem 4.7). Hence
$$\mathbb{I}^{*}_{\mathbb{S}} \subsetneq \mathbb{I}_{\mathbf{OP}} \ .$$

- Let $\mathbb{I}_{\mathbb{QL}}^{*}$ denote the set of QL-implications obtained from the triplet $(T_{\mathbf{M}}, S, N_S)$ where $T_{\mathbf{M}}(x,y) = \min(x,y)$, S is any t-conorm and N_S, the natural negation of S, is a strong negation and the pair (S, N_S) is such that $S(N_S(x), x) = 1,\ \ x \in [0,1]$. Then every $I \in \mathbb{I}_{\mathbb{QL}}^{*}$ satisfies (OP) ([5], Section 4.4). Hence
$$\mathbb{I}_{\mathbb{QL}}^{*} \subsetneq \mathbb{I}_{\mathbf{OP}}\ .$$

- Let $\mathbb{I}_{\mathbb{QL}}^{**}$ denote the set of QL-implications obtained from the triplet $(T_{\mathbf{M}}, S, N_S)$ where N_S is a strong negation and S is a right continuous t-conorm. Then every $I \in \mathbb{I}_{\mathbb{QL}}^{**}$ satisfies (OP) ([5], Section 4.4). Hence
$$\mathbb{I}_{\mathbb{QL}}^{**} \subsetneq \mathbb{I}_{\mathbf{OP}}\ .$$

For examples of fuzzy implications from other well-known families, viz., (U,N)-, RU-implications and the relationships among the properties they satisfy, please see, for instance, [4, 6] or the works of Bustince *et al.* [7, 8].

7. $\mathbb{F}^{T}_{\rightarrow_{\mathbf{OP}}}$ are Universal Approximators

Let us denote by $R^{T}_{\rightarrow_{\mathbf{OP}}}$ the fuzzy relation where the fuzzy implication $\longrightarrow$ is from $\mathbb{I}_{\mathbf{OP}}$, and the corresponding FRI by $\mathbb{F}^{T}_{\rightarrow_{\mathbf{OP}}} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^{T}_{\rightarrow_{\mathbf{OP}}}, d\right)$ where, $\mathcal{P}_X = \{A_i\}_{i=1}^{n} \subseteq \mathcal{F}(X)$ and $\mathcal{P}_Y = \{B_i\}_{i=1}^{n} \subseteq \mathcal{F}(Y)$.

Recall from (FRI-R-Singleton), for any $y \in Y$,

$$\begin{aligned} B'(y) &= R^{T}_{\rightarrow}(x_0, y) = T_{i=1}^{n}(A_i(x_0) \longrightarrow_{\mathbf{OP}} B_i(y)) \\ &= T\big(A_1(x_0) \longrightarrow_{\mathbf{OP}} B_1(y), A_2(x_0) \longrightarrow_{\mathbf{OP}} B_2(y), \ldots, A_n(x_0) \longrightarrow_{\mathbf{OP}} B_n(y)\big). \end{aligned} \tag{14}$$

Note that to get a final crisp output $y' \in Y$, we need to defuzzify the above $B' \in \mathcal{F}(Y)$ using d.

In this section, we show that, FRIs $\mathbb{F}^{T}_{\rightarrow_{\mathbf{OP}}} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^{T}_{\rightarrow_{\mathbf{OP}}}, \mathrm{MOM}\right)$ are universal approximators, i.e., the system function g of $\mathbb{F}^{T}_{\rightarrow_{\mathbf{OP}}}$, as defined in (9), can approximate any continuous function over a compact set to arbitrary accuracy. Moreover, we show that the system function g is continuous as discussed in [37].

Since the range of a continuous function over a closed and bounded interval is also a closed and bounded interval, in the following results, letting $X = [a,b]$ and if $h : X \to \mathbb{R}$ is any continuous function, we have $Y = h(X) = [c,d]$.

Theorem 7.1. *For any continuous function $h\colon [a,b] \to [c,d]$ over a closed and bounded interval and an arbitrary given $\varepsilon > 0$, there is an FRI $\mathbb{F}^{T}_{\rightarrow_{\mathbf{OP}}} = \left(\mathcal{P}_{[a,b]}, \mathcal{P}_{[c,d]}, R^{T}_{\rightarrow_{\mathbf{OP}}}, \mathrm{MOM}\right)$ with $\mathcal{P}_{[a,b]}$ and $\mathcal{P}_{[c,d]}$ being Ruspini partitions such that*

(i) *the system function g as defined in* (9) *is continuous on $[a,b]$, and*
(ii) $\max\limits_{x\in[a,b]} |h(x) - g(x)| < \varepsilon$.

Proof. Let $h\colon [a,b] \to [c,d]$ be any continuous function and let an $\varepsilon > 0$ be given. We present a procedural and stepwise proof of how to realise the FRI $\mathbb{F}^{T}_{\rightarrow_{\mathbf{OP}}}$ with the specified properties.

Step I : Choosing the points of normality

Since h is continuous over a closed interval $[a,b]$, h is uniformly continuous on $[a,b]$.
Thus for a given $\varepsilon > 0$ there exists $\delta > 0$ (depending on ε) such that, for all $w, w' \in [a,b]$,

$$|w - w'| < \delta \Longrightarrow |h(w) - h(w')| < \frac{\varepsilon}{2}\ .$$

Step I *(a)*: A Coarse Initial Partition

With the δ defined above and taking $l = 1 + \lceil \frac{b-a}{\delta} \rceil$ we now choose $w_i \in [a, b], i = 1, 2, \ldots l$, such that $|w_i - w_{i+1}| < \delta$, where $\lceil r \rceil$ is the integral value of the number r.

Let $z_i = h(w_i)$, the value h takes at the above chosen w_i, for $i = 1, 2, \ldots l$. We call these points w_i and z_i the points of normality on the input space and the output space, respectively.

***Step I** (b)*: Redundancy Removal and Reordering

Let us choose the distinct z_i's from the above and sort them in ascending order. Let $\sigma : \mathbb{N}_l \longrightarrow \mathbb{N}_k$ denote the above permuation map such that $z_i = u_{\sigma(i)}$, for $i = 1, 2, \ldots l$ and $u_j,\ j = 1, 2, \ldots, k$ are in ascending order.

***Step I** (c)*: Refinement of the input space partition:

Thus for each $i = 1, 2, \ldots, l$ we have $h(w_i) = z_i = u_{\sigma(i)}$. However, note that consecutive points of normality w_i, w_{i+1} in the input space need not be mapped to consecutive points of normality $u_{\sigma(i)}, u_{\sigma(i)+1}$ or $u_{\sigma(i)}, u_{\sigma(i)-1}$.

To ensure the above, we further refine the input space partition. To this end, we refine every sub-interval $[w_i, w_{i+1}]$, for $i = 1, 2, \ldots l - 1$ as follows. Note that $h(w_{i+1}) = u_{\sigma(i+1)}$.

Refinement Procedure:

For every $i = 1, 2, \ldots l - 1$ do the following:

(i) If $u_{\sigma(i+1)} = u_{\sigma(i)+1}$ or $u_{\sigma(i)-1}$ then we do nothing.

(ii) Let $u_{\sigma(i+1)} = u_{\sigma(i)+p}$, where $p \geq 2$. For every $u \in \{u_{\sigma(i)+1}, u_{\sigma(i)+2}, \ldots, u_{\sigma(i)+p-1}\}$ we find a point $v \in [w_i, w_{i+1}]$ such that $h(v) = u$. Note that the existence of such a $v \in [w_i, w_{i+1}]$ is guaranteed by the continuity - essentially the ontoness - of the function h. If $u = u_{\sigma(i)+q}$, for some $1 \leq q \leq p - 1$, then we denote the point v as $w_{i,i+1}^{(q)}$.

(iii) Similarly, let $u_{\sigma(i+1)} = u_{\sigma(i)-p}$, where $p \geq 2$. For every $u \in \{u_{\sigma(i)-1}, u_{\sigma(i)-2}, \ldots, u_{\sigma(i)-p+1}\}$ we find a $v \in [w_i, w_{i+1}]$ such that $h(v) = u$. Once again, if $u = u_{\sigma(i)-q}$, for some $1 \leq q \leq p - 1$, then we denote v as $w_{i,i+1}^{(q)}$.

***Step I** (d)*: Final Points of Normality:

Once the above process is done, we again rename the points of normality in the input space, viz., $w_{i,i+1}^{(q)}$ as $x_1, x_2, \ldots, x_n (n \geq l)$ and the $u_{\sigma(i)}$'s of the the output space as $y_1, y_2, \ldots y_k$.

For a graphical illustration of **Step I** given above, please see Section 7.1.

Step II : Construction of the Fuzzy Partitions - $\mathcal{P}_{[a,b]}, \mathcal{P}_{[c,d]}$

In the next step, we construct fuzzy sets on both the input and output spaces with the above obtained x_i's and y_j's as the points of normality, as given below.

***Step II** (a)*: Fuzzy Partition on the input space $\mathcal{P}_{[a,b]} = \{A_i\}_{i=1}^n$.

We construct n fuzzy sets such that

- Supp $A_i = (x_{i-1}, x_{i+1})$ for $i = 2, \ldots, n - 1$, while Supp $A_1 = [x_1, x_2)$ and Supp $A_n = (x_{n-1}, x_n]$,
- each A_i is normal at x_i, i.e., $A_i(x_i) = 1$,
- each A_i is a continuous convex fuzzy set, strictly increasing on $[x_{i-1}, x_i]$ and strictly decreasing on $[x_i, x_{i+1}]$.
- $\{A_i\}_{i=1}^n$ form a Ruspini partition (see. for instance (1)).

***Step II** (b)*: Fuzzy Partition on the output space $\mathcal{P}_{[c,d]} = \{C_j\}_{j=1}^k$.

We construct k fuzzy sets in a similar way as above, such that

- Supp$(C_j) = (y_{j-1}, y_{j+1})$ for $j = 2, \ldots, k - 1$, while Supp$(C_1) = [y_1, y_2)$ and Supp$(C_k) = (y_{k-1}, y_k]$,
- each C_j is normal at y_j, i.e., $C_j(y_j) = 1$,
- each C_j is a continuous convex fuzzy set, strictly increasing on $[y_{j-1}, y_j]$ and strictly decreasing on $[y_j, y_{j+1}]$.

- $\{C_j\}_{j=1}^{k}$ form a Ruspini partition.

Here obviously, $|y_j - y_{j-1}| < \frac{\varepsilon}{2}, \quad j = 1, 2, \ldots k$.

Step III: Construction of the rule base

We construct the rule base with n rules of the form:

$$\text{IF } \tilde{x} \text{ is } A_i \text{ THEN } \tilde{y} \text{ is } C_j\ , \quad i = 1, 2, \ldots n, \tag{15}$$

where the consequent C_j in the i-th rule is chosen such that $y_j = h(x_i)$, where x_i is the point at which A_i attains normality.

We can rewrite the rule base (15) as follows to confirm to the notations as in (4):

$$\text{IF } \tilde{x} \text{ is } A_i \text{ THEN } \tilde{y} \text{ is } B_i\ , \quad i = 1, 2, \ldots n\ . \tag{16}$$

Firstly, note that not all B_i's may be distinct. Further, since h is continuous, by the above assignment of the rules, we have that rules whose antecedents are adjacent also have adjacent consequents, i.e., for any $i = 1, 2, \ldots n-1$ we have Supp $B_i \cap$ Supp $B_{i+1} \neq \emptyset$.

In fact, the constructed rule base is also smooth [35], [36], [38].

Step IV : Approximation capability of the output

Let $x' \in [a, b]$ be the arbitrary given input. Clearly, $x' \in [x_m, x_{m+1}]$ for some $m \leq n-1$. Once again, by our construction, x' belongs to atmost two adjacent A_i's, and they are A_m, A_{m+1}. Thus, from (14),

$$\begin{aligned} B'(y) &= T\left[A_m(x') \longrightarrow_{\mathbf{OP}} B_m(y), A_{m+1}(x') \longrightarrow_{\mathbf{OP}} B_{m+1}(y)\right] \\ &= T\left[s_m \longrightarrow_{\mathbf{OP}} B_m(y), s_{m+1} \longrightarrow_{\mathbf{OP}} B_{m+1}(y)\right] \end{aligned}$$

where we introduce the notations $s_m = A_m(x')$ and $s_{m+1} = A_{m+1}(x')$ for better readability in the proofs. Note that since A_i's form a Ruspini partition, we have that $s_m + s_{m+1} = 1$. Further, note that by the construction of $\{A_i, B_i\}$, B_m, B_{m+1} are adjacent fuzzy sets.

Consider the kernel of B'. We choose the defuzzified output y' such that it belongs to Ker B'. In fact, as we show below, by the construction of $\{A_i, B_i\}$ we see that Ker B' is a singleton and this becomes the defuzzified output.

Since T is a t-norm, we know that $T(p, q) = 1$ if and only if $p = 1$ and $q = 1$. Further, note that $p \longrightarrow_{\mathbf{OP}} q = 1$ if and only if $p \leq q$ and $s_m + s_{m+1} = 1$ and hence we have

$$\begin{aligned} \text{Ker } B' &= \{y | B'(y) = 1\} \\ &= \{y | s_m \longrightarrow_{\mathbf{OP}} B_m(y) = 1\} \bigcap \\ &\qquad \{y | s_{m+1} \longrightarrow_{\mathbf{OP}} B_{m+1}(y) = 1\} \\ &= \{y | s_m \leq B_m(y)\} \bigcap \{y | s_{m+1} \leq B_{m+1}(y)\}\ . \end{aligned}$$

Let $\alpha_m = \inf\{\alpha | s_m \longrightarrow_{\mathbf{OP}} \alpha = 1\}$ and $\beta_{m+1} = \inf\{\beta | s_{m+1} \longrightarrow_{\mathbf{OP}} \beta = 1\}$. Since $\longrightarrow_{\mathbf{OP}}$ has (OP), clearly $\alpha_m = s_m$ and $\beta_{m+1} = s_{m+1}$.

By the continuity and convexity of B_m, B_{m+1} there exist $a_m, b_m, a_{m+1}, b_{m+1}$ such that $B_m(a_m) = B_m(b_m) = s_m$ and $B_{m+1}(a_{m+1}) = B_{m+1}(b_{m+1}) = s_{m+1}$. By the monotonicity of the implication in the second variable, for every $y \in [a_m, b_m]$ we have that $s_m \to B_m(y) = 1$ and for every $y \in [a_{m+1}, b_{m+1}]$ we have that $s_{m+1} \to B_{m+1}(y) = 1$. Thus,

$$\begin{aligned} \{y | s_m \leq B_m(y)\} &= [a_m, b_m]\ , \\ \{y | s_{m+1} \leq B_{m+1}(y)\} &= [a_{m+1}, b_{m+1}]\ , \quad \text{and} \\ \text{Ker } B' = \{y | B'(y) = 1\} &= [a_m, b_m] \bigcap [a_{m+1}, b_{m+1}]. \end{aligned}$$

Claim: Ker $B' = \{a_{m+1}\} = \{b_m\} \neq \emptyset$.

Firstly, note that for any $s_m \in [0,1]$ by the normality of B_m we have that $B_m(y_m) = 1$ and hence $y_m \in \{y \mid s_m \leq B_m(y)\} \Longrightarrow y_m \in [a_m, b_m] \neq \emptyset$. Similarly, $y_{m+1} \in [a_{m+1}, b_{m+1}] \neq \emptyset$. It suffices to show that $a_{m+1} \leq b_m$ from whence Ker $B' = [a_{m+1}, b_m]$.

Note that since $m < m+1$ and B_m, B_{m+1} are adjacent fuzzy sets, either $y_m < y_{m+1}$ or $y_m > y_{m+1}$. Without loss of generality, let us assume $y_m < y_{m+1}$. Now, from $a_{m+1} \in$ Supp B_{m+1} we have that $y_m \leq a_{m+1} \leq y_{m+1}$. Similarly, $y_m \leq b_m \leq y_{m+1}$. Hence, $y_m \leq a_{m+1}, b_m \leq y_{m+1}$. Since,

$$\begin{aligned} s_m + s_{m+1} = 1 &\Longrightarrow B_{m+1}(a_{m+1}) + B_m(b_m) = 1, \\ &\Longrightarrow B_{m+1}(a_{m+1}) = 1 - B_m(b_m), \\ &\Longrightarrow B_{m+1}(a_{m+1}) = B_{m+1}(b_m), \\ &\Longrightarrow b_m \in [a_{m+1}, b_{m+1}], \\ &\text{i.e., } a_{m+1} \leq b_m \ . \end{aligned}$$

Now, to see that $b_m = a_{m+1}$, note that since $\{B_i\}$ form a Ruspini partition and B_m, B_{m+1} are adjacent fuzzy sets, we have $B_{m+1}(a_{m+1}) = 1 - B_m(a_{m+1})$ and hence

$$B_m(a_{m+1}) = s_m = B_m(b_m). \tag{17}$$

Since $b_m, a_{m+1} \in$ Supp $B_m \cap$ Supp B_{m+1} on which both B_m, B_{m+1} are strictly monotonic (but of opposite types) we have that $b_m = a_{m+1}$.

Since d is the MOM defuzzification, we get that $g(x') = d(B') = a_{m+1} = b_m \in [y_m, y_{m+1}]$.

Claim: g is continuous on $[a, b]$.

Let us consider an $x' \in [a, b]$. Clearly, $x' \in [x_m, x_{m+1}]$ for some $1 \leq m < n$ and $g(x') = b_m \in [y_m, y_{m+1}]$.

To show that g is continuous at x', we need to show that for any given $\varepsilon > 0$, we can find a $\delta > 0$ such that, for any $x^* \in [a, b]$, whenever

$$|x^* - x'| < \delta \ \text{ then } \ |g(x^*) - g(x')| < \varepsilon. \tag{18}$$

Since B_m is strictly decreasing and continuous on $[y_m, y_{m+1}]$, we have that $B_m^{-1} : [0,1] \longrightarrow [y_m, y_{m+1}]$ exists. Thus from (17) we have $b_m = B_m^{-1}(s_m)$.

Further, B_m^{-1} is strictly decreasing and continuous on $[0,1]$. Hence, for any $\varepsilon_1 > 0$ there exists some $\delta_1 > 0$ such that for any $s_m^* \in [0,1]$,

$$|s_m^* - s_m| < \delta_1 \Longrightarrow |B_m^{-1}(s_m^*) - B_m^{-1}(s_m)| < \varepsilon_1. \tag{19}$$

Since $A_m : [x_m, x_{m+1}] \longrightarrow [0,1]$ is continuous, for any $\varepsilon_2 > 0$ there exists some $\delta_2 > 0$ such that

$$|x^* - x'| < \delta_2 \Longrightarrow |A_m(x^*) - A_m(x')| < \varepsilon_2. \tag{20}$$

Let $s_m^* = A_m(x^*)$. Then $g(x^*) = b_m^* \in [y_m, y_{m+1}]$ and

$$|s_m^* - s_m| = |A_m(x^*) - A_m(x')| \ , \text{ and} \tag{21}$$

$$|g(x^*) - g(x')| = |b_m^* - b_m| = |B_m^{-1}(s_m^*) - B_m^{-1}(s_m)| \ . \tag{22}$$

Now, let us set $\varepsilon_1 = \varepsilon$ and $\varepsilon_2 = \delta_1$. Then, for $\delta = \delta_2$, we have

$$\begin{aligned} |x^* - x'| < \delta &\Longrightarrow |A_m(x^*) - A_m(x')| < \varepsilon_2 \ , && \lceil\text{using (20)} \\ &\Longrightarrow |s_m^* - s_m| < \varepsilon_2 = \delta_1 \ , && \lceil\text{using (21)} \\ &\Longrightarrow |B_m^{-1}(s_m^*) - B_m^{-1}(s_m)| < \varepsilon_1 = \varepsilon \ , && \lceil\text{using (19)} \\ &\Longrightarrow |b_m^* - b_m| < \varepsilon \ , \\ &\Longrightarrow |g(x^*) - g(x')| < \varepsilon \ . && \lceil\text{using (22)} \end{aligned}$$

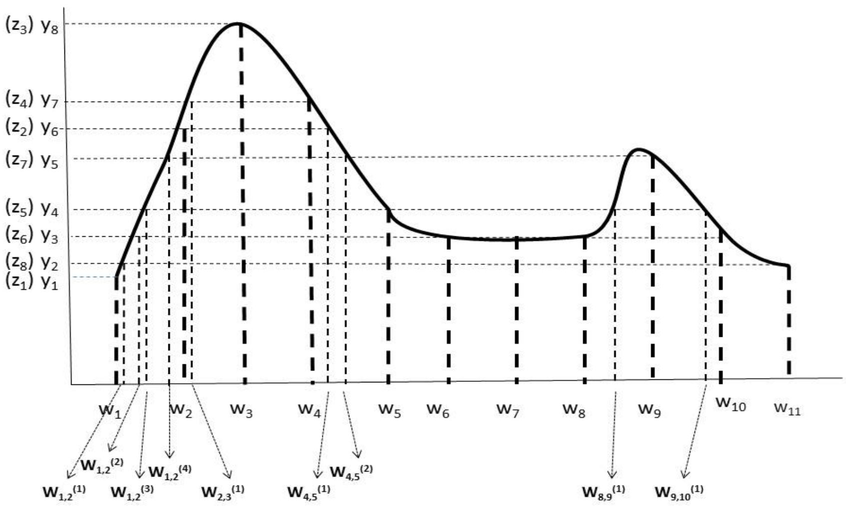


Figure 3: An Illustrative Example for **Step I** in the proof of Theorem 7.1. The intersection of the thin-dotted and thick-dotted lines with the x-axis give the points w_i and $w_{i,i+1}^{(q)}$, respectively.

Thus for any $\varepsilon > 0$, there exists a $\delta > 0$ such that, whenever $|x^* - x'| < \delta$ then $|g(x^*) - g(x')| < \varepsilon$, i.e., g is continuous on $[a, b]$.

Clearly, now,

$$|y_m - g(x')| < \frac{\varepsilon}{2} \quad \text{or} \quad |y_{m+1} - g(x')| < \frac{\varepsilon}{2} \ .$$

Without loss of generality, let $|y_m - g(x')| < \dfrac{\varepsilon}{2}$.

Further, since $x' \in [x_m, x_{m+1}]$ we have $|h(x') - y_m| < \dfrac{\varepsilon}{2}$. Putting them all together, we have

$$\begin{aligned} |g(x') - h(x')| &\leq |g(x') - y_m| + |y_m - h(x')| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

Since x' is arbitrary we have, $\max\limits_{x \in [a,b]} |h(x) - g(x)| < \varepsilon$.

7.1. Some Illustrative Remarks on the proof of Theorem 7.1

In this section we illustrate some steps of the proof of the Theorem 7.1 through figures and examples for better understanding.

Step I : Choosing the points of normality

***Step I** (a)*: A Coarse Initial Partition

In Fig. 3, the points $w_1, w_2, \ldots, w_{11}$ and the points $z_1, z_2, \ldots z_8$ (in paranthesis) are the points of normality in the input and the output spaces, respectively.

***Step I** (b)*: Redundancy Removal and Reordering *By rearranging the z_i's in ascending order and renaming them we obtain: $u_1 = z_1 < u_2 = z_8 < u_3 = z_6 < u_4 = z_5 < u_5 = z_7 < u_6 = z_2 < u_7 = z_4 < u_8 = z_3$.*

***Step I** (c)*: Refinement of the input space partition: *In Fig. 3, $h(w_1) = u_1$ and $h(w_2) = u_6$. Thus for the consecutive points w_1 and w_2 the function values are u_1 and u_6, which are not consecutive.*

Refinement Procedure:

From Fig. 3, it can be seen that we have inserted points $w_{1,2}^{(1)}, w_{1,2}^{(2)}, w_{1,2}^{(3)}, w_{1,2}^{(4)} \in [w_1, w_2]$. Proceeding similarly, the following sub-intervals, shown in Fig. 3, have been refined: $[w_2, w_3], [w_4, w_5], [w_8, w_9]$ and $[w_9, w_{10}]$.

***Step II** (a)*: Fuzzy Partition on the input space $\mathcal{P}_{[a,b]} = \{A_i\}_{i=1}^{n}$.

For instance, let each of the A_i's ($i = 2, \ldots, n-1$) be a triangular fuzzy set, (i.e, each A_i is linear and strictly increasing on $[x_{i-1}, x_i]$, each A_i is linear and strictly decreasing on $[x_i, x_{i+1}]$), let A_1 be right-half-triangular, (i.e., A_1 is linear and strictly decreasing on $[x_1, x_2)$) and let A_n be left-half-triangular, (i.e., A_n is linear and strictly increasing on $(x_{n-1}, x_n]$). Further, let all the A_i's attain normality at x_i. Then, clearly, the fuzzy partition $\{A_i\}_{i=1}^{n}$ of the input space $[a, b]$ is a Ruspini partition and each of the A_i's is continuous, convex, of finite support and $A_i(x_i) = 1$.

8. $\mathbb{F}^{T}_{\to_{\mathbf{D1}}}$ are Universal Approximators

While in the previous section, we dealt with fuzzy implications satisfying (OP), this class of fuzzy implications is rather limited. In this section, we consider those positive implications whose natural negations are Gödel negation.

Let us denote by $R^{T}_{\to_{\mathbf{D_1}}}$ the fuzzy relation where the fuzzy implication $\longrightarrow$ is from $\mathbb{I}^{+}_{N_{\mathbf{D1}}}$ and the corresponding FRI by $\mathbb{F}^{T}_{\to_{\mathbf{D1}}} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^{T}_{\to_{\mathbf{D_1}}}, d\right)$ where, $\mathcal{P}_X = \{A_i\}$ and $\mathcal{P}_Y = \{B_i\}$.

Once again, recall that from (FRI-R-Singleton), with $R = R^{T}_{\to_{\mathbf{D_1}}}$ for any $y \in Y$, we have

$$\begin{aligned} B'(y) &= R^{T}_{\to_{\mathbf{D_1}}}(x_0, y) = T_{i=1}^{n}(A_i(x_0) \longrightarrow_{\mathbf{D_1}} B_i(y)) \\ &= T\big(A_1(x_0) \longrightarrow_{\mathbf{D_1}} B_1(y), A_2(x_0) \longrightarrow_{\mathbf{D_1}} B_2(y), \ldots, A_n(x_0) \longrightarrow_{\mathbf{D_1}} B_n(y)\big). \end{aligned} \tag{23}$$

We now show that the FRIs $\mathbb{F}^{T}_{\to_{\mathbf{D1}}} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^{T}_{\to_{\mathbf{D1}}}, \mathrm{MOM}\right)$ are universal approximators, i.e., they can approximate any continuous function over a compact set to arbitrary accuracy.

Theorem 8.1. *For any continuous function $h\colon [a, b] \to [c, d]$ over a closed interval and an arbitrary given $\varepsilon > 0$, there is an FRI $\mathbb{F}^{T}_{\to_{\mathbf{D_1}}} = \left(\mathcal{P}_{[a,b]}, \mathcal{P}_{[c,d]}, R^{T}_{\to_{\mathbf{D_1}}}, \mathrm{MOM}\right)$ with $\mathcal{P}_{[a,b]}$ and $\mathcal{P}_{[c,d]}$ being Ruspini partitions such that the system function g approximates h uniformly, i.e., $\max\limits_{x\in[a,b]} |h(x) - g(x)| < \varepsilon$.*

Proof. Once again the proof is given in four steps. **Steps I–III** dealing with the construction of the input and output partitions and the rule base are done in exactly the same way as in **Steps I–III** of the proof of Theorem 7.1.

Step IV: Approximation capability of the output

Once again, let $x' \in [a, b]$ be the arbitrary given input. Clearly, $x' \in [x_m, x_{m+1}]$ for some $m \le l - 1$. Once again, by our construction, x' belongs to A_m, A_{m+1}. Thus, from (23),

$$\begin{aligned} B'(y) &= T\left[A_m(x') \longrightarrow_{\mathbf{D_1}} B_m(y), A_{m+1}(x') \longrightarrow_{\mathbf{D_1}} B_{m+1}(y)\right] \\ &= T\left[s_m \longrightarrow_{\mathbf{D_1}} B_m(y), s_{m+1} \longrightarrow_{\mathbf{D_1}} B_{m+1}(y)\right] = T\left[B'_m(y), B'_{m+1}(y)\right] \end{aligned}$$

where $s_m = A_m(x')$ and $s_{m+1} = A_{m+1}(x')$. Note that since A_i's form a Ruspini partition, we have that $s_m + s_{m+1} = 1$.

Now since $\longrightarrow_{\mathbf{D_1}}$ is positive and such that $x \longrightarrow_{\mathbf{D_1}} 0 = 0$ for any $x \in (0, 1]$, we have from Remark 6.7 that the supports of both the modified fuzzy sets B'_m, B'_{m+1} are the same as those of B_m, B_{m+1}, i.e., Supp B'_m=Supp B_m and Supp B'_{m+1} = Supp B_{m+1}. Hence,

$$\begin{aligned} \text{Supp } B' &= \text{Supp } B'_m \cap \text{Supp } B'_{m+1} \\ &= \text{Supp } B_m \cap \text{Supp } B_{m+1} \\ &= \text{Supp } (B_m \cap B_{m+1}) = [y_m, y_{m+1}]\ . \end{aligned} \tag{24}$$

Now, since (24) holds, we have

$$g(x') = \mathrm{MOM}(B') \in \mathrm{Supp}\ (B_m \cap B_{m+1}) = [y_m, y_{m+1}]\ .$$

Thus, $|y_m - g(x')| < \dfrac{\varepsilon}{2}$ and $|y_{m+1} - g(x')| < \dfrac{\varepsilon}{2}$. Once again, since $x' \in [x_m, x_{m+1}]$ we have $|h(x') - y_m| < \dfrac{\varepsilon}{2}$. Putting them all together, we have

$$|g(x') - h(x')| \leq |g(x') - y_m| + |y_m - h(x')| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} < \varepsilon.$$

Since $x' \in [a, b]$ is arbitrary, we have that $\max_{x \in [a,b]} |h(x) - g(x)| < \varepsilon$.

9. Illustrative Examples

In this section we illustrate our results through some examples. We consider 3 functions one each from the following three types or classes of functions, viz., those that are (i) purely monotonic, (ii) mixed monotonic and symmetric, and (iii) mixed monotonic and asymmetric. We then approximate these functions by the FRIs $\mathbb{F}^T_{\to\mathbf{OP}}$ or $\mathbb{F}^T_{\to\mathbf{D1}}$ as proposed and constructed in Sections 7 and 8.

We consider the Mean of Maxima defuzzification and the input and output space partitions are constructed as detailed in Section 7. However, we consider different fuzzy implication operators I coming from both the classes, viz., $\mathbb{I}_{\mathbf{OP}}$ and $\mathbb{I}^{+}_{N_{\mathbf{D1}}}$, in the examples.

In the given figures, the original functions $h(x)$ are shown in thick lines, the approximating system functions $g(x)$ in thin lines and the bounds $h(x) - \varepsilon$ and $h(x) + \varepsilon$ are plotted using dotted $--$ lines.

Example 9.1. *Let us consider the function*

$$h(x) = \ln(x), \quad x \in [2, 7]\ ,$$

which is strictly increasing on the interval $[2, 7]$ *and let* $\varepsilon = 0.1$*. According to the proposed construction we obtain* 50 *rules, since* $\delta = \varepsilon = 0.1$*. We approximate* h *using the FRI* $\mathbb{F}^T_{\to\mathbf{OP}}$*, where the implication operator employed in the relation* $R^T_{\to\mathbf{OP}}$ *is the Rescher implication* $I_{\mathbf{RS}} \in \mathbb{I}_{\mathbf{OP}} \setminus \mathbb{I}^{+}_{N_{\mathbf{D1}}}$ *(see Table 1), which satisfies* (OP) *and its natural negation* $N_{I_{\mathbf{RS}}} = N_{\mathbf{D1}}$*, but* $I_{\mathbf{RS}}$ *is not positive. The function* h *and its approximation* g *are shown in Fig. 4.*

Example 9.2. *Let us consider the function*

$$h(x) = \sin(x), \quad x \in [-2\pi, 2\pi]\ ,$$

which is mixed monotonic and symmetric on the interval $[-2\pi, 2\pi]$*. However, note that it is piecewise strictly increasing or decreasing. Let* $\varepsilon = 0.1$*. According to the proposed construction we obtain* 130 *rules. We approximate* h *using the FRI* $\mathbb{F}^T_{\to\mathbf{OP}}$*, where the implication operator employed in the relation* $R^T_{\to\mathbf{OP}}$ *is the Łukasiewicz implication* $I_{\mathbf{LK}} \in \mathbb{I}_{\mathbf{OP}} \setminus \mathbb{I}^{+}_{N_{\mathbf{D1}}}$ *(see Table 1) which satisfies* (OP) *and is positive, but* $N_{I_{\mathbf{LK}}} \neq N_{\mathbf{D1}}$*. The function* h *and its approximation* g *are shown in Fig. 5.*

Example 9.3. *Let us consider the function*

$$h(x) = -x^4 + 2x^2 - x, \quad x \in [-2, 2]\ ,$$

which is both mixed monotonic and asymetric on the interval $[-2, 2]$*. Let* $\varepsilon = 2$*.*

It is clear from the proof of Theorem 7.1 that the δ *obtained for a given* ε *is extremely conservative. Thus, in this case we would get a* $\delta = 0.0025$*. However, we have assumed* $\delta = 0.25$ *and proceeded to verify the approximation capability. According to the proposed construction, with* $\delta = 0.25$*, we obtain* 20 *rules by* ***Step I****(a) of Theorem 7.1, which are then refined to* 38 *rules by* ***Step I****(b) and* ***Step I****(c) of Theorem 7.1. We approximate* h *using the FRI* $\mathbb{F}^T_{\to\mathbf{D}_1}$*, where the implication operator employed in the relation* $R^T_{\to\mathbf{D}_1}$ *is*

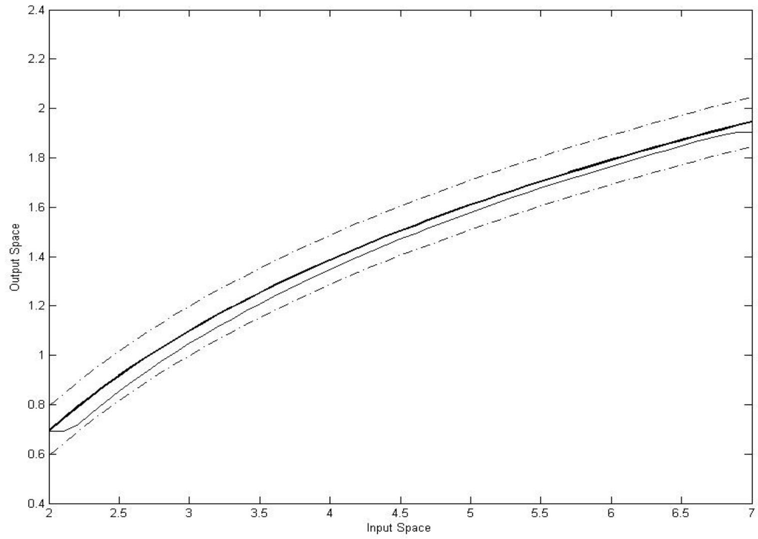


Figure 4: The natural logarithm function $h(x) = \ln x$ approximated within $\varepsilon = 0.1$ bound over $[2, 7]$

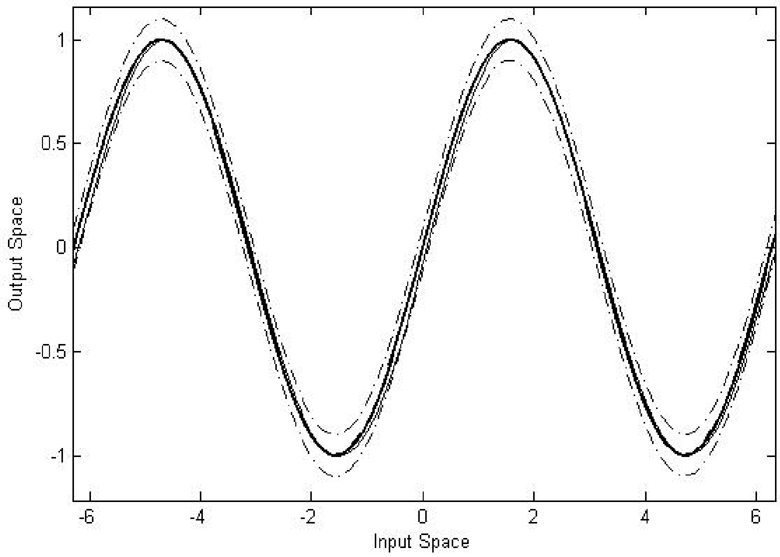
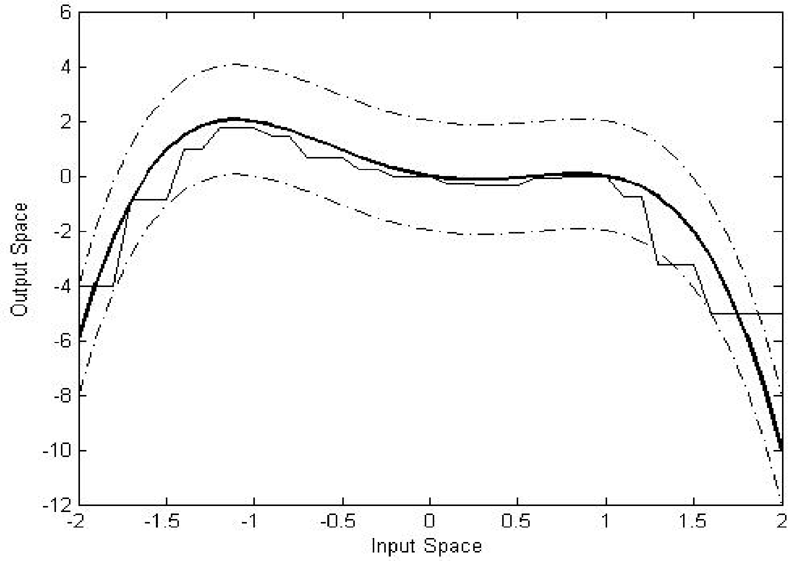


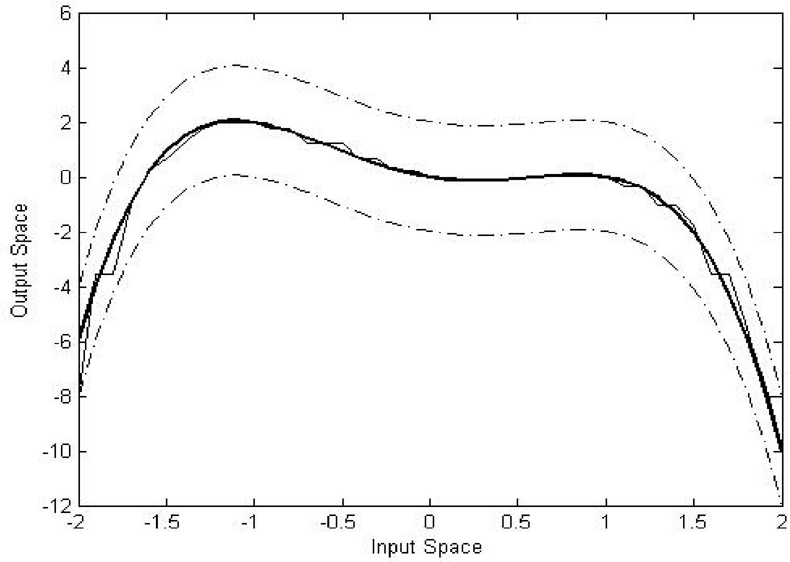
Figure 5: The function $h(x) = \sin(x)$ approximated within $\varepsilon = 0.1$ bound over $[-2\pi, 2\pi]$.

the Yager implication $I_{\mathbf{YG}} \in \mathbb{I}^{+}_{N_{\mathbf{D1}}} \setminus \mathbb{I}_{\mathbf{OP}}$ *(see Table 1) which does not satisfy* (OP)*, but is both positive and* $N_{I_{\mathbf{YG}}} = N_{\mathbf{D1}}$.

The approximated function is shown in Fig. 6. Fig. 6(a) gives the plot of the approximator g that was obtained from the original rule base with 20 *rules that were obtained before the refinement of the input space, while Fig. 6(b) gives the plot of the system function g that was obtained by employing the refined rule base with* 38 *rules.*



(a) Using 20 rules obtained from the unrefined partition



(b) Using 38 rules obtained from the refined partition

Figure 6: A 4th degree polynomial $h(x) = -x^4 + 2x^2 - x$ approximated over $[-2, 2]$ within $\varepsilon = 2$ bound.

10. Concluding Remarks

In this work, we provide a constructive proof of the universal approximation properties of FRIs when implicative rules are employed, i.e., where the operation between the antecedents and consequents is taken as a fuzzy implication.

We show that FRIs of the form $\mathbb{F}^T_{\rightarrow} = \left(\mathcal{P}_X, \mathcal{P}_Y, R^T_{\rightarrow}, d\right)$ are universal approximators. The proof is general enough for a large class of fuzzy implications I - without making any assumptions about the form or representation of the considered fuzzy implications - and is valid for any continuous function, not necessarily monotonic and the partitions used are of the Ruspini type.

From ***Step I****(a)* of the proof of Theorem 7.1, note that the number of rules generated from the proposed construction in Section 7 is dependent both on the function and the ε value given but not on the fuzzy implication employed in $\mathbb{F}^T_{\rightarrow}$. It is also clear from the illustrated examples that even a coarser partition of the input space than what is proposed can still approximate the given function within the bounds, i.e., even with a bigger δ we can still get the same ε approximation. Of course, the δ itself can be adapted depending on the prior knowledge of the slope of the given function to be approximated.

In the literature, it is typical to consider an R-implication obtained from a left-continuous t-norm in the fuzzy relation $\hat{R}_{\rightarrow}$ (see (7)) to model a rule base. However, in this work, the R that is considered is much more general, i.e., the fuzzy implications considered come from a much larger set than just the R-implications obtained from a left-continuous t-norm, which satisfy (OP) and hence ensure the coherence of the implicative model of the rule base. Further, the concept of *weak coherence* proposed in this work played an important role in enlarging the class of fuzzy implications that can be considered for implicative models. Section 6.4 lists some families of fuzzy implications that are admissible under this context.

Recently, in [25], it was shown that the Bandler-Kohout Subproduct inference (BKS-R) employing the Yager's families of fuzzy implications [42], namely, f- and g-implications, and the rule base modeled by the following relation,

$$R^T_{\rightarrow}(x,y) = \hat{R}_Y(x,y) = \bigwedge_{i=1}^{n} (A_i(x) \longrightarrow_Y B_i(y)), \tag{25}$$

where $\longrightarrow_Y$ is once again either an f- or a g-implication, i.e., $\longrightarrow_Y$ belongs to $\mathbb{I}_{\mathbb{F}} \cup \mathbb{I}_{\mathbb{G}}$ (see Section 6.4), have many desirable properties namely, interpolativity, continuity and robustness. Note that all f- and g-implications are positive and those that belong to $\mathbb{I}_{\mathbb{F},\infty} \cup \mathbb{I}_{\mathbb{G}}$ are such that their natural negations are the Gödel negation. This work clearly demonstrates that such FRIs are also universal approximators.

Thus, we believe that these results are very much applicable in most of the practical and desirable contexts as discussed in these works [17, 40, 39] and show that a much larger class of fuzzy implications other than what is typically considered in the literature can be employed meaningfully in FRIs based on implicative models.

Acknowledgements

We would like to thank Mr. Avijit Ashe, who helped us with the implementation of the FRIs proposed in Theorems 7.1 and 8.1.

References

[1] Baczyński, M., Jayaram, B., 2007. Yager's classes of fuzzy implications: some properties and intersections. Kybernetika 43, 157–182.
[2] Baczyński, M., Jayaram, B., 2008a. Fuzzy Implications. volume 231 of *Studies in Fuzziness and Soft Computing*. Springer.
[3] Baczyński, M., Jayaram, B., 2008b. (S,N)- and R-implications: A state-of-the-art survey. Fuzzy Sets and Systems 159, 1836–1859.
[4] Baczyński, M., Jayaram, B., 2009. (U,N)-implications and their characterizations. Fuzzy Sets and Systems 160, 2049 – 2062.
[5] Baczyński, M., Jayaram, B., 2010. QL-implications: Some properties and intersections. Fuzzy Sets and Systems 161, 158–188.
[6] Baczyński, M., Jayaram, B., 2011. Intersections between some families of (U,N)- and RU-implications. Fuzzy Sets and Systems 167, 30 – 44.
[7] Bustince, H., Burillo, P., Soria, F., 2003. Automorphisms, negations and implication operators. Fuzzy Sets and Systems 134, 209 – 229.
[8] Bustince, H., Fernandez, J., Sanz, J., Baczyński, M., Mesiar, R., 2013. Construction of strong equality index from implication operators. Fuzzy Sets and Systems 211, 15 – 33.
[9] Butnariu, D., 1983. Additive fuzzy measures and integrals I. Journal of Mathematical Analysis and Applications 93, 436 – 452.

[10] Castro, J., 1995. Fuzzy logic controllers are universal approximators. IEEE Trans. on Systems, Man and Cybernetics 25, 629–635.

[11] Castro, J., Delgado, M., 1996. Fuzzy systems with defuzzification are universal approximators. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, 26, 149–152.

[12] Coufal, D., 2006. Coherence of radial implicative fuzzy systems, in: IEEE International Conference on Fuzzy Systems, pp. 307–314.

[13] Cross, V., Sudkamp, T., 1993. Fuzzy implication and compatibility modification, in: IEEE International Conference on Fuzzy Systems, pp. 219–224 vol.1.

[14] De Baets, B., Mesiar, R., 1998. T-partitions. Fuzzy Sets and Systems 97, 211 – 223.

[15] Driankov, D., Hellendoorn, H., Reinfrank, M., 1996. An introduction to fuzzy control (2nd ed.). Springer-Verlag, London, UK.

[16] Dubois, D., Prade, H., 1996. What are fuzzy rules and how to use them. Fuzzy Sets and Systems 84, 169 – 185.

[17] Dubois, D., Prade, H., Ughetto, L., 1997. Checking the coherence and redundancy of fuzzy knowledge bases. IEEE Transactions on Fuzzy Systems 5, 398–417.

[18] Höhle, U., 1998. Many-valued equalities, singletons and fuzzy partitions. Soft Computing 2, 134–140.

[19] Jayaram, B., 2007. Yager's new class of implications J_f and some classical tautologies. Information Sciences 177, 930 – 946.

[20] Jayaram, B., 2008. Rule reduction for efficient inferencing in similarity based reasoning. International Journal of Approximate Reasoning 48, 156–173.

[21] Klement, E.P., Mesiar, R., Pap, E., 2000. Triangular Norms. volume 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht.

[22] Li, Y.M., Shi, Z.K., Li, Z.H., 2002a. Approximation theory of fuzzy systems based upon genuine many-valued implications: MIMO cases. Fuzzy Sets and Systems. 130, 159–174.

[23] Li, Y.M., Shi, Z.K., Li, Z.H., 2002b. Approximation theory of fuzzy systems based upon genuine many-valued implications: SISO cases. Fuzzy Sets and Systems. 130, 147–157.

[24] Magrez, P., Smets, P., 1989. Fuzzy modus ponens: A new model suitable for applications in knowledge-based systems. International Journal of Intelligent Systems 4, 181–200.

[25] Mandal, S., Jayaram, B., . Bandler-Kohout subproduct with Yager's classes of fuzzy implications. IEEE Trans. Fuz Sys. URL: `http://dx.doi.org/10.1109/TFUZZ.2013.2260551`.

[26] Markechová, D., 1992. The entropy of fuzzy dynamical systems and generators. Fuzzy Sets and Systems 48, 351 – 363.

[27] Montes, S., Couso, I., Gil, P., 2003. Fuzzy δ - ϵ-partitions. Information Sciences 152, 267 – 285.

[28] Pedrycz, W., 1985. Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data. Fuzzy Sets and Systems. 16, 163–175.

[29] Perfilieva, I., Kreinovich, V., 2002. A new universal approximation result for fuzzy systems, which reflects CNF–DNF duality. International Journal of Intelligent Systems 17, 1121–1130.

[30] Roychowdhury, S., Pedrycz, W., 2001. A survey of defuzzification strategies. International Journal of Intelligent Systems 16, 679–695.

[31] Sugeno, M., Kang, G.T., 1988. Structure identification of fuzzy model. Fuzzy Sets and Systems. 28, 15–33.

[32] Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics SMC-15, 116 –132.

[33] Tikk, D., Kóczy, L.T., Gedeon, T.D., 2003. A survey on universal approximation and its limits in soft computing techniques. International Journal of Approximate Reasoning 33, 185–202.

[34] Turksen, I., Zhong, Z., 1988. An approximate analogical reasoning approach based on similarity measures. IEEE Transactions on Systems, Man and Cybernetics 18, 1049 –1056.

[35] Van Broekhoven, E., De Baets, B., 2008. Monotone Mamdani–Assilian models under mean of maxima defuzzification. Fuzzy Sets and Systems 159, 2819 – 2844.

[36] Van Broekhoven, E., De Baets, B., 2009. Only smooth rule bases can generate monotone Mamdani–Assilian models under center-of-gravity defuzzification. IEEE Trans. on Fuzzy Systems 17, 1157–1174.

[37] Štěpnička, M., Bodenhofer, U., Daňková, M., Novák, V., 2010a. Continuity issues of the implicational interpretation of fuzzy rules. Fuzzy Sets and Systems 161, 1959 – 1972.

[38] Štěpnička, M., De Baets, B., 2013. Implication-based models of monotone fuzzy rule bases. Fuzzy Sets and Systems 232, 134 – 155. doi:`http://dx.doi.org/10.1016/j.fss.2013.07.019`.

[39] Štěpnička, M., De Baets, B., Nosková, L., 2010b. Arithmetic fuzzy models. IEEE Trans. on Fuzzy Systems 18, 1058 –1069.

[40] Štěpnička, M., Jayaram, B., 2010. On the suitability of the Bandler-Kohout subproduct as an inference mechanism. IEEE Trans. on Fuzzy Systems 18, 285–298.

[41] Wang, L.X., Mendel, J., 1992. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. IEEE Transactions on Neural Networks 3, 807 –814.

[42] Yager, R.R., 2004. On some new classes of implication operators and their role in approximate reasoning. Information Sciences 167, 193 – 216.

[43] Yager, R.R., Kreinovich, V., 2003. Universal approximation theorem for uninorm-based fuzzy systems modeling. Fuzzy Sets and Systems 140, 331–339.

[44] Zadeh, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics SMC-3, 28 –44.

[45] Zadeh, L.A., 1975. The concept of a linguistic variable and its application to approximate reasoning-III. Information Sciences 9, 43–80.

[46] Zeng, X.J., Singh, M., 1994. Approximation theory of fuzzy systems-SISO case. IEEE Transactions on Fuzzy Systems 2, 162–176.