# A Fuzzy Inference Methodology Based on the Fuzzification of Set Inclusion

Chris Cornelis and Etienne E. Kerre

Department of Mathematics and Computer Science, Ghent University
Fuzziness and Uncertainty Modelling Research Unit
Krijgslaan 281 (S9), B-9000 Gent, Belgium

**Abstract.** Nowadays, people start to accept fuzzy rule–based systems as flexible and convenient tools to solve a myriad of ill–defined but otherwise (for humans) straightforward tasks such as controlling fluid levels in a reactor, automatical lens focussing in cameras and adjusting an aircraft's navigation to the change of winds and so on. Contrary to the intuition often seen as the feeding ground of fuzzy rule–based systems—namely, that they realize an extension of the Modus Ponens (MP) rule of inference to an environment with more than two truth–values—most actual applications rely at the base level on common interpolation techniques or similarity assessments to simulate the process of "calculating with words" perceived at the user level. It is doubtful whether these somewhat opportunistic approaches will perform well when more challenging requirements (e.g. aspects of logical consistency; incorporation of varying facets of uncertainty) are imposed in order to implement a successful artificial reasoning unit. Therefore, in this paper, starting from the notion of a fuzzy restriction (i.e. the basic building block of our rule–based system) we list some elementary consistency requirements that a fuzzy inference system should satisfy. Subsequently we describe a reasoning methodology based on a measure of fulfilment of the antecedent clause of an if–then rule. Inclusion–based approximate reasoning, as we coined it in [7], outperforms the traditional scheme based on the Compositional Rule of Inference (CRI) in terms of both complexity and of logical soundness. In terms of semantics it also offers a better solution to the implementation of analogical reasoning than similarity measures are able to do.

**Keywords:** fuzzy expert systems, fuzzy inclusion measures, compositional rule of inference, analogical reasoning

## 1 Introduction

*Fuzzy rules* have proven to be a very accurate and effective tool for specifying how a given system should operate. For instance, man, to solve everyday tasks, routinely moves through a series of actions, guided by the information his senses provide him with. Whether the obtained information is exact (as in the ideal situation of an abstract game) or ill–defined (as in most real–world problems), good solution strategies are generally arrived at by trying to match[1] perceptions of the actual world state to previously encountered ones and by consulting available knowledge so as to come up with the most profitable action. A convenient way of summing up such knowledge is by a series of if–then rules of the form:

If $X$ satisfies restriction $A$ then $Y$ satisfies restriction $B$

In this expression, $X$ and $Y$ are system variables that could represent just about anything from measurable physical quantities like speed or weight up to qualitative judgments about, e.g., the certainty or belief that a particular action can be safely executed. Each variable is assumed to take values in its own domain (which could be the set of real numbers, a set of selected linguistic labels, ...). Clauses are then formed by characterizing these variables in terms of values that they may assume; a process that comes down to limiting, in a flexible way, the possibility that a variable takes on a given value, as in the clause:

"John is old"

In the next section we will give a formal definition of this knowledge representation formalism.

Inference, then, is defined as a procedure for deducing new facts out of existing ones on the basis of formal deduction rules. Classical paradigms devised for this purpose, such as two–valued propositional and predicate logic, exhibit some important drawbacks (lack of expressivity, high computational complexity) that make them unsuitable for application in automated deduction systems. To allow for a higher degree of flexibility and expressivity, Zadeh in 1973 introduced a formalism called approximate reasoning to cope with problems which are too complex for exact solution but which do not require a high degree of precision. [22]

Section 2 will define the central inference pattern for reasoning with fuzzy restrictions—known as Generalized Modus Ponens (GMP)—and will also identify some elementary properties of inference regarding logical consistency. Next, we revisit two classical inference strategies, the first based on the Compositional Rule of Inference (CRI), the second using similarity measures, and discuss some of their drawbacks and advantages. Particularly, we will stress that the popular class of fuzzy controllers, whose importance is evidenced by the large influx in our economy of intelligent "fuzzy–logic enhanced" electronic devices like washing machines, rice cookers, etc. actually bypass the inference step, and constitute little more than sophisticated interpolation devices cleverly creating an illusion of "doing mathematics with words".

Inclusion–based reasoning, first introduced in [7], is dealt with in section 3. Some new results regarding the fuzzification of set inclusion obtained in [8]

[1] In case of incomplete or uncertain information we are likely to try and draw an *analogy* between a prototype and an observation.

enhance our original treatment of the concept; moreover our results are generalized to cover a system of parallel rules rather than one rule in isolation. A theorem linking our approach with the traditional CRI–based scheme is repeated from [7] and supplemented with some considerations about the consistency of a batch of rules. Finally, section 4 offers a brief summary and some options for future research.

## 2 Classical Inference Strategies

In this section we explain the concept of a fuzzy or elastic restriction and show how it can be used in reasoning processes.

We assume that a system consists of a number of input and output variables, and that each variable $X$ can take values in its own universe $U$. In some cases we might not be able to express its value precisely, which is quite common when we are dealing with linguistic information. For instance, assume $X$ is a real quantity representing a person's age and that we have been provided with the information that "$X$ is about 25". We can then assign to each element of the universe a degree between 0 and 1, expressing the possibility that $X$ takes on precisely this value, and the result would be a graph (called a possibility distribution) looking like the one depicted in figure 1.

**Fig. 1.** Possibility distribution of $X$

Evidently, this kind of flexible or fuzzy restriction on the values that $X$ may assume can be interpreted as the membership function of a fuzzy set $A$. For reasons of consistency, we demand that there be at least one value of U which is completely possible for $X$. Before going on we introduce some additional terminology: by the support of a fuzzy set $A$, denoted $Supp(A)$, we mean the crisp set of elements of the universe that belong to a degree strictly higher than zero to $A$:

$$Supp(A) = \{u \in U | A(u) > 0\} . \tag{1}$$

The kernel of $A$, $Ker(A)$ is the set of elements that belong fully to $A$:

$$Ker(A) = \{u \in U | A(u) = 1\} . \tag{2}$$

A fuzzy set $A$ for which $Ker(A) \neq \emptyset$ holds is called normalized. In the context of inference patterns a fuzzy restriction "$X$ is $A$" is also termed a *fuzzy fact*, while a *fuzzy rule* is built up by combining two fuzzy facts "$X$ is $A$" and "$Y$ is $B$" into the joint clause

IF $X$ is $A$ THEN $Y$ is $B$

From a logical perspective, it is interesting to see how people are able to combine such imprecise information efficiently in a Modus Ponens–like fashion to allow for inferences of the following kind:

| | |
|---|---|
| IF the bath water is "too hot" | THEN I'm likely to get burnt |
| bath water is "really rather hot" | |
| | I'm quite likely to get burnt |

The technique used above is in fact less restrictive than the actual Modus Ponens from propositional logic since it does not require the observed fact ("really rather hot") and the antecedent of the rule ("too hot") to coincide to yield a meaningful conclusion. The need emerges for a flexible, qualitative scale of measuring to what extent the antecedent is fulfilled, on the basis of which we could obtain an approximate idea (stated under the form of another fuzzy restriction) of the value of the consequent variable.

With the introduction of a calculus of fuzzy restrictions [22], Zadeh paved the way towards a reasoning scheme called Generalized Modus Ponens (GMP) to systematize deductions like the example we presented:

**Definition 1. (Generalized Modus Ponens, GMP)** Let $X$ and $Y$ be variables assuming values in $U$, resp. $V$. Consider then a fuzzy rule "IF $X$ is $A$, THEN $Y$ is $B$" and a fuzzy fact (or observation) "$X$ is $A'$" ($A, A' \in \mathcal{F}(U), B \in \mathcal{F}(V)$, where $\mathcal{F}(U)$ denotes the class of fuzzy sets in $U$). The GMP allows then deduction of a fuzzy fact "$Y$ is $B'$", with $B' \in \mathcal{F}(V)$.

Expressing this under the form of an inference scheme, we get:

$$\frac{\begin{array}{r}\text{IF } X \text{ is } A, \text{ THEN } Y \text{ is } B\\ X \text{ is } A'\end{array}}{Y \text{ is } B'}$$

The above pattern does not state what the fuzzy restriction $B'$ should be when $A, A'$ and $B$ are given; indeed, it is not a computational procedure. Before turning our attention to such procedures, it is worthwhile considering for a moment which criteria we like GMP to satisfy. Four really important ones are listed below:

| | |
|---|---|
| A.1 $B \subseteq B'$ | (nothing better than $B$ can be inferred) |
| A.2 $A'_1 \subseteq A'_2 \Rightarrow B'_1 \subseteq B'_2$ | (monotonicity) |
| A.3 $A' = A \Rightarrow B' = B$ | (compatibility with modus ponens) |
| A.4 $A' \subseteq A \Rightarrow B' = B$ | (fulfilment of $A$ implies fulfilment of $B$) |

The first three are all standard in the approximate reasoning literature (see e.g. [1] [11]); A.4, which is, strictly speaking, superfluous as it is a direct consequence of A.1, A.2 and A.3 combined, paraphrases the following intuition: whenever the restriction $A'$ on $X$ is at least as specific as $A$, the outcome of the inference should be exactly $B$. This criterion becomes more clear when we interpret the subset sign here as "$A'$ is completely fulfilled by $A$", in which case of course the rule has full applicability.

Zadeh suggested to model an if–then rule as a fuzzy relation $R$ (a fuzzy set on $U \times V$) and to apply the Compositional Rule of Inference (CRI), a convenient and intuitively sound mechanism for calculating with fuzzy restrictions introduced in [22], to yield an inference about $Y$. We repeat the definition from [6]:

**Definition 2. (Compositional Rule of Inference, CRI)** Let $X$ and $Y$ be variables taking values in $U$, resp. $V$. Furthermore consider fuzzy facts "$X$ is $A$" and "$X$ and $Y$ are $R$", where $A \in \mathcal{F}(U), R \in \mathcal{F}(U \times V)$ ($R$ is a fuzzy relation between $U$ and $V$). The CRI allows us to infer the fuzzy fact: "$Y$ is $R \circ_T A$", in which the fuzzy composition of $R$ and $A$ by the $t$–norm[2] $T$, denoted $R \circ_T A$ is defined as, for $v \in V$:

$$R \circ_T A(v) = \sup_{u \in U} T(A(u), R(u,v))\,. \tag{3}$$

Expressing this under the form of an inference scheme, we get:

$$\begin{array}{l} X \text{ is } A \\ X \text{ and } Y \text{ are } R \\ \hline Y \text{ is } R \circ_T A \end{array}$$

The motivation behind this rule stems from a calculus of fuzzy restrictions where a join of fuzzy facts, e.g. one about $X$ and one about $Y$, is performed by looking for the least specific restriction on the tuple $(X, Y)$ (i.e. putting the least conditions on them), that is consistent with each of the original restrictions.

Applying this rule to the GMP, for every $v \in V$ we have to evaluate the following formula:

$$B'(v) = \sup_{u \in U} T(A'(u), R(u,v))\,. \tag{4}$$

We will refer to the above approach as CRI–GMP, i.e. a realization[3] of GMP by CRI.

Since Zadeh's pioneering work, many researchers have sought for efficient realizations of this approximate inference scheme. It should be clear that different choices of $R$ and $T$ in the CRI–GMP scheme yield systems with

[2] A $t$–norm is any symmetric, associative, increasing $[0,1]^2 \to [0,1]$ mapping $T$ satisfying $T(1,x) = x$ for every $x \in [0,1]$

[3] By "realization", we mean any computational procedure unambiguously defining the output in terms of the inputs

substantially different characteristics. If we want definition 2 to be a realization of the GMP that preserves the consistency criteria A.1 through A.4, it can be verified that $R$ should be a relational representation of a fuzzy implicator, an extension of the classical implication operator:

**Definition 3. (Fuzzy implicator)** [14] A fuzzy implicator is any $[0,1]^2 \rightarrow [0,1]$ mapping $\mathcal{I}$ for which the restriction to $\{0,1\}^2$ coincides with classical implication: $\mathcal{I}(0,0)=1$, $\mathcal{I}(1,0)=0$, $\mathcal{I}(0,1)=1$, $\mathcal{I}(1,1)=1$. Moreover, $\mathcal{I}$ should satisfy the following monotonicity criteria:

$$(\forall y \in [0,1])(\forall (x,x') \in [0,1]^2)(x \leq x' \Rightarrow \mathcal{I}(x,y) \geq \mathcal{I}(x',y))$$
$$(\forall x \in [0,1])(\forall (y,y') \in [0,1]^2)(y \leq y' \Rightarrow \mathcal{I}(x,y) \leq \mathcal{I}(x,y'))$$

A multitude of fuzzifications of the implication operator to model $R$ have been proposed in the literature. Table 1 lists some important classes of fuzzy implicators.[4] After choosing a fuzzy implicator $\mathcal{I}$, we put $R(u,v) = \mathcal{I}(A(u),B(v))$ for all $(u,v) \in U \times V$.

**Table 1.** Fuzzy implicators on the unit interval ($(x,y) \in [0,1]^2$)

| Symbol | Name | Definition | Comment |
|---|---|---|---|
| $\mathcal{I}_S$ | $\mathcal{S}$–implicator | $\mathcal{I}_S(x,y) = S(1-x,y)$ | $S$ a $t$–conorm |
| $\mathcal{I}_T$ | $\mathcal{R}$–implicator (residual implicator) | $\mathcal{I}_T(x,y) = \sup\{\gamma \in [0,1] \vert T(x,\gamma) \leq y\}$ | $T$ a $t$–norm |
| $\mathcal{I}_{T,S}$ | QL–implicator (quantum logic implicator) | $\mathcal{I}_{T,S}(x,y) = S(1-x,T(x,y))$ | $T$ a $t$–norm, $S$ a $t$–conorm |

The suitability of a given $(T,\mathcal{I})$ pair to implement the CRI–GMP can be evaluated with respect to the listed criteria. Extensive studies have been carried out on this issue (see e.g. [11]); the following theorem shows that for a given *continuous* $t$–norm, the fuzzy implicator can always be chosen so that A.1 through A.4 simultaneously hold.

**Theorem 1.** *Let $T$ be a continuous $t$–norm. The CRI–GMP based on the $(T,\mathcal{I}_T)$ pair satisfies A.1–A.4.*

*Proof.*

- From $T(A(u),B(v)) \leq B(v))$ for any $(u,v) \in U \times V$ we easily find that $\mathcal{I}_T(A(u),B(v)) = \sup\{\gamma \in [0,1] \vert T(A(u),\gamma) \leq B(v)\} \geq B(v)$. Hence,

[4] A $t$–conorm is any symmetric, associative, increasing $[0,1]^2 \rightarrow [0,1]$ mapping $S$ satisfying $S(0,x) = x$ for every $x \in [0,1]$

$B'(v) = \sup_{u \in U} T(A'(u), \mathcal{I}_T(A(u), B(v))) \geq \sup_{u \in U} T(A'(u), B(v))$. Since we assumed that $A'$ is normalized, this expression is bounded below by $T(1, B(v)) = B(v)$, so A.1 holds.

- By taking into account the monotonicity of $t$–norms, it is easily verified that A.2 also holds.
- The continuity of $T$ allows the following deduction, for $(u, v) \in U \times V$:

$$\begin{aligned} T(A(u), \mathcal{I}_T(A(u), B(v))) &= T(A(u), \sup\{\gamma \in [0,1] | T(A(u), \gamma) \leq B(v)\}) \\ &= \sup\{T(A(u), \gamma) | \gamma \in [0,1] \wedge T(A(u), \gamma) \leq B(v)\} \\ &\leq B(v) \end{aligned}$$

In other words $B'(v) \leq B(v)$. But since also $B'(v) \geq B(v)$, we obtain $B' = B$ and hence A.3.

Obviously, A.4 is also satisfied as a consequence of A.1, A.2 and A.3 combined. This completes the proof. □

One particularly unfortunate aspect of this use of the CRI is its high complexity. In general, for finite universes $U$ and $V$ so that $|U| = m$ and $|V| = n$, an inference requires $O(mn)$ operations. Mainly for this reason, researchers have explored other ways of performing fuzzy inference. In fuzzy control, the common strategy chooses $T$ as the minimum operator and defines the relation $R$, for $u$ in $U$ and $v$ in $V$, as:

$$R(u, v) = \min(A(u), B(v)) . \tag{5}$$

It is easily verified that the following computationally very efficient formula emerges ($v \in V$):

$$B'(v) = \min\left(\sup_{u \in U}(A'(u), A(u)), B(v)\right) . \tag{6}$$

Note that this choice allows for a significant reduction in complexity: a deduction now only requires $O(m+n)$ operations. On the other hand it can be verified that A.1 (hence also A.4) cannot be maintained under this approach. In fact, as Klawonn and Novak remarked in [13], the above calculation rule is not a logical inference, since no logical implication[5] is inside and thus no modus ponens proceeds. The latter strategy is sometimes referred to as the conjunction–based model of CRI–GMP, while the former (which does involve logical inference) is called the implication–based model. Klawonn and Novak showed that the conjunction–based model, when applied to a collection of parallel fuzzy rules, amounts, at the base level, to simple interpolation.

[5] Notwithstanding this, some researchers stick to the ill–chosen terminology of Mamdami "implicator" for minimum.

Another line of research (see e.g. [3] [16] [21]) is called analogical reasoning and relies on a paradigm based on analogy to cut back on the number of calculations: when $A'$ gets close to $A$, $B'$ should likewise become closer to B. In other words, the fuzzy rules are treated as typical instances of the system's behaviour, and the objective is try to draw as close as possible an analogy between the observed situation and these prototypes. This principle is used in similarity–based reasoning, where it gives way to the following realization of the GMP:

$$\begin{array}{l} \text{IF } X \text{ is } A, \text{ THEN } Y \text{ is } B \\ \quad X \text{ is } A' \\ \hline \qquad\qquad\qquad\qquad Y \text{ is } f(\mathcal{S}(A',A),B) \end{array}$$

First, $A'$ is compared to $A$ by means of a similarity measure $\mathcal{S}$, which will yield a number in the interval [0,1], on the basis of which a mapping $f$ will modify the consequent into the outcome $B'$. Similarity is not a uniquely defined notion; care should be taken when adopting such or such interpretation for use in a given application; nevertheless $\mathcal{S}$ is normally assumed to be at least reflexive and symmetric, so as to act intuitively correct as an indicator of the resemblance between fuzzy sets. One example of such a measure is given here, where we investigate the intersection of $A'$ and $A$ defined by the minimum, and we look for the element with the highest membership degree in it and return this value as the similarity of $A'$ and $A$:

$$\mathcal{S}(A',A) = \sup_{u\in U} \min(A'(u), A(u))\,. \tag{7}$$

As an example of a modification mapping $f$, we quote the following formula from [4], which introduces a level of *uncertainty* proportional to $1-\alpha$, thus making inference results easy to interpret: the higher this value gets, the more the original $B$ is "flooded" by letting the minimal membership grade in $B'$ for every $v$ in $V$ become at least $1-\alpha$:

$$f(\alpha,B)(v) = \max(1-\alpha, B(v))\,. \tag{8}$$

Since the similarity between $A'$ and $A$ needs to be calculated only once, the overall complexity of this scheme is again $O(|U|+|V|)$.

Again, with respect to the consistency requirements, it will be difficult if not impossible to come up with a modification mapping $f$ and a similarity measure $\mathcal{S}$ that satisfy all 4 of them. For the purpose of modelling analogy, symmetry is actually both counterintuitive and harmful! Counterintuitive, because we compare an observation $A'$ to a reference $A$ and not the other way around; harmful, because imposing symmetry inevitably clashes with the soundness condition A.4 (from $A' \subseteq A$ infer $B' = B$), and renders inference absolutely useless (just imagine a *symmetrical* measure $\mathcal{S}$ that satisfies $\mathcal{S}(A,B) = 1$ if $A \subseteq B$). While the idea of analogical reasoning is in itself very useful, its execution so far has often been subject to opportunistic flaws that have put it, from a logical perspective, on the wrong track.

In the next section, we will propose an alternative strategy based on a fuzzification of crisp inclusion, and argue why it is better than the existing methods.

## 3 Inclusion–Based Approach

Instead of concentrating on the similarity of fuzzy sets, it makes more sense to consider the fulfilment of one fuzzy restriction by another, that is: to check whether the observation $A'$ is a subset of the antecedent $A$ of the fuzzy rule. Bearing in mind the close relationship between fulfilment and inclusion, we might capture this behaviour provided we can somehow measure the degree of inclusion of $A'$ into $A$.

Indeed, if we have such a measure (say $Inc(A', A)$), we can use it to transform the consequent fuzzy set $B$ into an appropriate $B'$. Schematically, this amounts to the following:

$$\begin{array}{l} \textbf{IF } X \text{ is } A \ \textbf{ THEN } Y \text{ is } B \\ \quad X \text{ is } A' \\ \hline \qquad\qquad\qquad\qquad Y \text{ is } f(Inc(A', A), B) \end{array}$$

with $f$ again a given modification mapping. Good candidates for the $(f, Inc)$ pair will preferably be such that A.1 through A.4 hold with as little extra conditions added as possible. In addition, we would like to have $Inc(B', B) = Inc(A', A)$, in order that a kind of symmetry between the fulfilment of $B'$ by $B$ and that of $A'$ by $A$ is respected. In the next subsections, we will consider the following three problems in details: the definition of suitable inclusion grades, and the description of an inclusion–based reasoning algorithm when a) only one if–then rule and b) a collection of parallel if–then rules are involved.

### 3.1 Fuzzification of Set Inclusion

Zadeh, in his seminal 1965 paper, was the first to propose a definition for the inclusion of one fuzzy set into another. It reads:

$$A \subseteq B \iff (\forall u \in U)(A(u) \leq B(u)) . \tag{9}$$

This rigid definition unfortunately does not do justice to the true spirit of fuzzy set theory: we may want to talk about a fuzzy set being "more or less" a subset of another one, and for this reason researchers have set out to define alternative indicators of the inclusion of one fuzzy set into another: i.e., $\mathcal{F}(U) \times \mathcal{F}(U) \to [0, 1]$ mappings $Inc$, such that the value $Inc(A, B)$ indicates to what extent $A$ is included into $B$. Of course we need to constrain the admissible class of indicators; an axiom scheme like the one proposed by Sinha and Dougherty [15] serves this purpose well.

**Definition 4. (Sinha–Dougherty Axioms)** Let $Inc$ be an $\mathcal{F}(U)\times\mathcal{F}(U) \rightarrow [0,1]$ mapping, and $A, B$ and $C$ fuzzy sets in a given universe $U$. The Sinha–Dougherty axioms imposed on $Inc$ are as follows:

**Axiom 1** $Inc(A,B) = 1 \iff A \subseteq B$ (in Zadeh's sense)
**Axiom 2** $Inc(A,B) = 0 \iff Ker(A) \cap co\ Supp(B) \neq \emptyset$,
**Axiom 3** $B \subseteq C \Rightarrow Inc(A,B) \leq Inc(A,C)$, i.e. the indicator has increasing second partial mappings
**Axiom 4** $B \subseteq C \Rightarrow Inc(C,A) \leq Inc(B,A)$, i.e. the indicator has decreasing first partial mappings
**Axiom 5** $Inc(A,B) = Inc(S(A), S(B))$
where $S$ is a $\mathcal{F}(U) \rightarrow \mathcal{F}(U)$ mapping defined by, for every $u \in U$, $S(A)(u) = A(s(u))$, $s$ denoting an $U \rightarrow U$ mapping
**Axiom 6** $Inc(A,B) = Inc(coB, coA)$
**Axiom 7** $Inc(B \cup C, A) = min(Inc(B,A), Inc(C,A))$
**Axiom 8** $Inc(A, B \cap C) = min(Inc(A,B), Inc(A,C))$

The second axiom might at first glance seem harsh (e.g. Wilmott [18] and Young [20] preferred to leave it out in favour of more compensating operators), but as Sinha and Dougherty [15] proved, it is indispensible if we want $Inc$ to be a faithful extension of the classical inclusion, that is, $Inc(A,B) \in \{0,1\}$ if $A$ and $B$ are crisp sets. The original version included a ninth axiom, $Inc(A, B \cup C) \geq \max(Inc(A,B),\ Inc(A,C))$. Frago [10] indicated that it is redundant because, as can easily be verified, it is equivalent to axiom 3.

Starting from a very general expression depending on four functional parameters for such an indicator, Sinha and Dougherty in [15] proposed conditions they claimed to be necessary and sufficient to satisfy the axioms. In [8] we revisited and corrected their findings, allowing for a simpler and more consistent framework for the axiomatic characterization of inclusion grades for fuzzy sets. For our purposes, the following theorem is of particular importance:

**Theorem 2.** *Let $U$ be a finite universe and $A$, $B$ fuzzy sets in $U$. When $Inc$ is defined as*

$$Inc(A,B) = \inf_{u \in U} \psi(A(u), B(u))\,. \tag{10}$$

*the conditions for the $[0,1]^2 \rightarrow [0,1]$ mapping $\psi$ listed in table 2 are necessary and sufficient to satisfy the S–D axioms 1 through 8.*

Table 2 conditions make it clear that any suitable $\psi$ will also be a fuzzy implicator. This is not surprising, given the well–known relationship between implication and inclusion in classical set theory: $A$ is a subset of $B$ ($A$, $B$ crisp sets defined in a universe $U$) if and only if

$$(\forall u \in U)(u \in A \Rightarrow u \in B)\,. \tag{11}$$

**Table 2.** Necessary and sufficient conditions for $\psi$ to satisfy the 8 S–D axioms

| | |
|---|---|
| I1 | $(\forall x_1, x_2, y \in [0,1])(x_1 \leq x_2 \Rightarrow \psi(x_1, y) \geq \psi(x_2, y))$ |
| I2 | $(\forall x, y_1, y_2 \in [0,1])(y_1 \leq y_2 \Rightarrow \psi(x, y_1) \leq \psi(x, y_2))$ |
| I3 | $(\forall x, y \in [0,1])(\psi(x, y) = \psi(1-y, 1-x))$ |
| I4 | $(\forall x, y \in [0,1])(x \leq y \iff \psi(x, y) = 1)$ |
| I5 | $(\forall x, y \in [0,1])(x = 1 \wedge y = 0 \iff \psi(x, y) = 0)$ |

On the other hand, not any fuzzy implicator satisfies I1–I5. In [8] we outlined a class of mappings, generalizations of the Łukasiewicz implicator, which all share table 2 properties:

**Definition 5. (Generalized Łukasiewicz implicator)** Every implicator $\mathcal{I}$ defined as, for $x$ and $y$ in $[0,1]$,

$$\mathcal{I}(x, y) = \min(1, \lambda(x) + \lambda(1-y)) \,. \tag{12}$$

where $\lambda$ is a strictly decreasing $[0,1] \to [0,1]$ mapping satisfying $\lambda(0) = 1$, $\lambda(1) = 0$ and

$$(\forall x, y \in [0,1])(x \leq y \iff \lambda(x) + \lambda(1-y) \geq 1) \,. \tag{13}$$

is called a generalized Łukasiewicz implicator.

Note that the restrictions I1–I5 are all accounted for by this definition. I1 and I2 are standard implicator requirements; I3 is easily seen to be satisfied as well; I4 is equivalent to condition (13); I5 requires that $\lambda$ strictly decrease.

To summarize, from now on we will be working with the following shape of an inclusion indicator $Inc$, for $A', A \in \mathcal{F}(U)$:

$$Inc_\lambda(A', A) = \inf_{u \in U} \min(1, \lambda(A'(u)) + \lambda(1 - A(u))) \,. \tag{14}$$

with $\lambda$ defined as in definition 5.

### 3.2 Inclusion–Based Reasoning with One Fuzzy Rule

We first consider the simple case of a system with only one rule. An inclusion measure will allow us to evaluate $\alpha = Inc_\lambda(A', A)$. This degree, in turn, will be used to modify the consequent $B$ of the considered rule into a suitable output $B'$, i.e. $B'(v) = f(\alpha, B)(v)$ for all $v$ in $V$ with $f$ a modification mapping of our choice.

To comply with condition A.1, it is clear that $f(\alpha, B)(v) \geq B(v)$. On the other hand, since $Inc_\lambda$ satisfies the monotonicity axiom 4, $f$ ought to be increasing w.r.t. its first argument to fulfil A.2. Lastly, to have A.3 and A.4 it is mandatory that $f(1, B) = B$, whatever $B \in \mathcal{F}(V)$.

As we have seen, the need for modification mappings also arises in similarity–based reasoning, where instead $\alpha$ will result from a similarity measurement, so we can "borrow", so to speak, some of the work which has been done in that field.

Several modification mappings serve our cause; deciding which one to choose depends largely on the application at hand. Nevertheless, in a situation where we would like the inference result to be in accordance somehow with the output of a given CRI–GMP system, one mapping might be considered more eligible than the next one. We have noticed that there exists a link between fuzzy inclusion and fuzzy implicators, so it really comes as no surprise that the behaviour of our approach can be closely linked to that of the CRI based on particular $t$–norm/implicator pairs. Indeed, in [7], we showed that for a residuated implicator generated by a continuous $t$–norm $T$, the following theorem and its important corollary hold:

**Theorem 3.** *Let $T$ be a continuous $t$–norm. If $B'$ represents the result obtained with CRI–GMP based on the $(T,\mathcal{I}_T)$ pair, i.e. for all $v \in V$*

$$B'(v) = \sup_{u\in U} T(A'(u), \mathcal{I}_T(A(u), B(v)))\,. \tag{15}$$

*and the inclusion measure $Inc_{\mathcal{I}_T}$ is defined as*[6]*, for $A', A \in \mathcal{F}(U)$:*

$$Inc_{\mathcal{I}_T}(A', A) = \inf_{u\in U} \mathcal{I}_T(A'(u), A(u))\,. \tag{16}$$

*then*

$$Inc_{\mathcal{I}_T}(B', B) \geq Inc_{\mathcal{I}_T}(A', A)\,. \tag{17}$$

*Additionally, if $(\forall\alpha \in [0,1])(\exists v \in V)(B(v) = \alpha)$, then*

$$Inc_{\mathcal{I}_T}(B', B) = Inc_{\mathcal{I}_T}(A', A)\,. \tag{18}$$

**Corollary 1.** *For every $v \in V$, the inference result $B'(v)$ obtained with CRI–GMP based on the $(T,\mathcal{I}_T)$ pair, where $T$ is a continuous $t$–norm, is bounded above by the expression $\mathcal{I}_T(Inc_{\mathcal{I}_T}(A', A), B(v))$.*

In effect, this shows that if we put $f(\alpha, B)(v) = \mathcal{I}_T(\alpha, B(v))$ for every $v$ in $V$, a conclusion entailed by our algorithm is a superset (not necessarily a proper one) of the according CRI–GMP result, which can be regarded as a justification of its soundness: indeed, when we replace the output of the CRI–GMP by a less specific fuzzy set, the corresponding constraint on the output variable $Y$ will likewise be less strong, since every value of the universe will be assigned at least as high a possibility degree by our strategy as by the original CRI–GMP inference mechanism. This is illustrated in figure 2.

[6] Note that the class introduced in equation (14) consists of specific instances of this pattern.

**Fig. 2.** Example of the link between the inclusion–based (*straight line*) and CRI–GMP (*dotted line*) inference result

### 3.3 Inclusion–Based Reasoning with Parallel Fuzzy Rules

For any realistic application, a single rule will not suffice to describe the relationship between the relevant system variables adequately. Therefore it makes sense to consider blocks of parallel rules like the generic rule base below:[7]

**IF** $X$ is $A_1$ **THEN** $Y$ is $B_1$
**IF** $X$ is $A_2$ **THEN** $Y$ is $B_2$
...
**IF** $X$ is $A_n$ **THEN** $Y$ is $B_n$

where of course we assume that for $i \neq j$, $A_i \neq A_j$.

In order to perform inference on the basis of a block of fuzzy rules, we need a mechanism to somehow execute the rules in parallel. For CRI–GMP, two widespread strategies which have complementary behaviour exist. [12] We review them briefly.

The first approach is called First Infer Then Aggregate (FITA): it entails a conclusion for each rule in isolation and then aggregates these results to the final system outcome. The other available option, First Aggregate Then Infer (FATI), will aggregate the fuzzy rules into one central rule that is subsequently used for CRI–GMP inference. Below we list the formula for CRI–GMP in each case, for $v \in V$:

$$FITA : B'(v) = \oplus_{i=1}^{n}\left(\sup_{u\in U} T(A'(u), R_i(u,v)\right)$$
$$FATI : B'(v) = \sup_{u\in U} T(A'(u), \oplus_{i=1}^{n} R_i(u,v))$$

where $\oplus$ is a mapping called aggregation operator [12] satisfying

1. $\oplus\underbrace{(0,...,0)}_{n \text{ times}} = 0$ and $\oplus\underbrace{(1,...,1)}_{n \text{ times}} = 1$ (border conditions)
2. For $(a_1,...,a_n),(b_1,...,b_n) \in [0,1]^n$ we have: (monotonicity)

$$(\forall i \in \{1,...,n\})(a_i \leq b_i) \Rightarrow \oplus(a_1,...,a_n) \leq \oplus(b_1,...,b_n)$$

[7] In some cases, both antecedent and consequent may involve a multitude of variables interacting in various ways, like in the rule:

**IF** $X$ is much bigger than $Y$ **THEN** either $Z$ or $U$ should be slightly reduced

Such a rule can still be adapted to our framework by treating the variables in the antecedent, resp. consequent, as a single compound variable taking values in a compound universe.

3. $\oplus$ is a continuous mapping

There is no obvious way to tailor FATI to the needs of inclusion–based reasoning since the fuzzy relation $R$ is never explicitly used. We can, on the other hand, adapt FITA in a meaningful way to obtain the following formula:

$$B'(v) = \oplus_{i=1}^{n} f(Inc_\lambda(A', A_i), B_i(v))\,. \tag{19}$$

What now remains is the choice of the aggregation operator; again, some considerations about the CRI–GMP can guide us here. In practice, for reasons of simplicity, system developers choose minimum as aggregation operator.[8] We briefly motivate the use of min as FITA aggregation operator: suppose the observation $A'$ and the antecedent $A_i$ of the $i^{\text{th}}$ rule are disjunct, i.e. $Supp(A') \cap Supp(A_i) = \emptyset$, then there exists $u^* \in U$ such that $A_i(u^*) = 0$ and $A'(u^*) = 1$ (since $Ker(A') \neq \emptyset$ by assumption). For every $v \in V$ we find, using the inclusion indicator shape introduced in equation (14):

$$\begin{aligned} Inc_\lambda(A', A_i) &= \inf_{u \in U} \min(1, \lambda(A'(u)) + \lambda(1 - A_i(u))) \\ &\leq \min(1, \lambda(A'(u^*)) + \lambda(1 - A(u^*)) = 0 \end{aligned}$$

Very often the modification mapping $f$ is chosen such that $f(0, B)(v) = 1$, for all $v \in V$, in order that the conclusion obtained with rule $i$ equals the universe $V$ of $Y$; in other words, it doesn't allow us to infer anything and its effects should not be taken into account when drawing a general conclusion. The aggregation operator must be able to cancel its effect. Minimum is therefore an obvious candidate. Unfortunately, min raises some complications with respect to the important criteria of coherency and consistency.

We call an inference strategy *coherent* if for every collection of parallel fuzzy rules, when the observation exactly matches the antecedent of one of the rules, the inference outcome equals the consequent of this rule. This is intuitively very acceptable, but in practise it is hard to realize (with CRI–GMP) because of the varying degrees of influence of different rules of the rule base on the final result. One option to avoid this anomaly is to detect beforehand whether the observation indeed matches one of the antecedents. Unfortunately this approach is unworkable because of two reasons:

- Comparing two fuzzy sets element by element is a computationally costly operation, especially if the universe involved is very large.
- Continuity of the system's outcome would suffer, since a very small variation in the observation $A'$ can give way to a substantially different inference result.

[8] That is, when the implication–based model of the CRI–GMP, to which our approach is closely linked, is used. For more details we refer to [12].

Buckley and Hayashi [5] describe an interpolation–based strategy using a weighted sum of membership degrees to have FITA aggregation coherent[9]; in other words, given the weight vector $(\mu_1, ..., \mu_n)$ the aggregation operator $\oplus$ is defined as, for $(x_1, ..., x_n) \in [0,1]^n$, as:

$$\oplus(x_1, ..., x_n) = \sum_{i=1}^{n} \mu_i x_i \,. \tag{20}$$

We briefly review this approach: a metric is used to assess the distance $D_i$ between the observation and the antecedent of the $i^{\text{th}}$ rule, and the distance $D_{ij}$ between antecedent $i$ and antecedent $j$. Then $\lambda_j$, for $j = 1, ..., n$ is defined as

$$\lambda_j = \min_{i=1, i \neq j}^{n} D_{ij} \,. \tag{21}$$

$\lambda_j > 0$ since all antecedents are assumed to be different. Next they define $\mu'_i$, for $i = 1, ..., n$, as:

$$\mu'_i = \begin{cases} \frac{\lambda_i - D_i}{\lambda_i} & \text{if } 0 \leq D_i \leq \lambda_i \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Unless $\delta = \sum_{i=1}^{n} \mu'_i = 0$ (in which case an unrestricted output should be produced), the scaled weight factors are calculated as follows:

$$\mu_i = \frac{\mu'_i}{\delta} \,. \tag{23}$$

In other words, when $B'_i$ represents the output obtained from the $i^{\text{th}}$ rule, Buckley and Hayashi's overall system output $B'$ is given by, for $v \in V$:

$$B'(v) = \sum_{i=1}^{n} \mu_i B'_i(v) \,. \tag{24}$$

They proved that the resulting inference strategy is coherent provided criterion A.3 (compatibility with modus ponens) holds for individual rules. We could directly apply Buckley and Hayashi's procedure to aggregate the individual results obtained with inclusion–based reasoning with one rule, but this solution is not very elegant since it relies on distances (and thus on similarities), which is not in the spirit of our proposed ideas.

It makes more sense to use a (scaled version of) the inclusion degrees $Inc(A', A_i)$ as weights. Indeed, define, for $i = 1, ..., n$:

$$\mu'_i = Inc(A', A_i) \,. \tag{25}$$

[9] Their ideas were directed at CRI–GMP, but they also apply in our case.

and again put $\delta = \sum_{i=1}^{n} \mu'_i$ (when $\delta = 0$, return $V$ as the system output) to calculate $\mu_i = \frac{\mu'_i}{\delta}$.

This aggregation procedure is not coherent as such, since from $Inc_\lambda(A', A_i) = 1$ for a given $i$ we cannot infer that $Inc_\lambda(A', A_j) = 0$ for $j \neq i$. If however the rule base satisfies the supplementary condition that, for $i, j$ in $\{1, ..., n\}$:

$$Supp(A_i) \cap Ker(A_j) = \emptyset \,. \tag{26}$$

then, as soon as $A' = A_i$ (hence $Inc_\lambda(A', A_i) = 1$), only one weight factor is different from zero; since $B'_i = B_i$ by A.3 the inference process is coherent.

Another characteristic of an inference strategy is called *consistency.* To be consistent, a system may only return valid fuzzy restrictions, i.e. normalized fuzzy sets. Subnormalization occurs when several (conflicting) fuzzy rules want to associate incompatible results with the output variable. [19] This can be particularly troublesome when results from one inference are used for further deductions (chaining of rules). It can be easily verified that, with weighted sum aggregation as presented above, the overall output need not necessarily be normalized as soon as more than one weight factor is strictly positive. In the remainder of this section, we will present a robust inference strategy preserving both coherency and consistency. It is called rule preselection and is due to Dvořák. [9]

The basic idea behind rule preselection is to extract from the rule base, for every given observation, only that fuzzy rule which is best in accordance with that fact. It is assumed that this rule provides us with sufficient information about what the conclusion should look like. Even though one might argue that the balancing behaviour of a "genuine" aggregation strategy (i.e. taking into account all partial results) is disposed of in this way, we gain considerably in speed and still don't violate coherency and consistency; that is, provided we choose our rule selection mechanism carefully. Dvořák proposes to use similarity degrees (selecting the rule whose antecedent best resembles the observation) for this purpose, but again, guided by our considerations about analogical reasoning, we prefer inclusion degrees.

To summarize, the algorithm for multiple rule inclusion–based reasoning with rule preselection is stated below.[10]

1. For $i = 1, ..., n$, calculate
$$\alpha_i = Inc(A', A_i) \,. \tag{27}$$
2. Choose rule $l$ with the highest value of $\alpha_l$

[10] Only one dilemma remains: what to do when, unlikely but not impossibly, the maximum inclusion degree is attained several times? Dvořák suggests to use an auxiliary criterion (a distance, or another inclusion measure, for example) to settle the case.

3. Compute the conclusion $B'$ by means of, for $v \in V$:

$$B'(v) = f(\alpha_l, B_l)(v)\,. \tag{28}$$

## 4 Conclusion

Several authors have pursued the strategy of analogical reasoning in their quest to model human behaviour in various cognitive tasks, such as classification and decision making. A panoply of similarity measures have been reviewed in the literature as a means to draw analogies between situations (fuzzy sets). For our purposes, the symmetry property of similarity measures is actually both counterintuitive and harmful! Counterintuitive, because we compare an observation $A'$ to a reference $A$ and not the other way round; harmful, because imposing symmetry inevitably clashes with the soundness condition A.4.

In this chapter, we redefined the semantics of analogical reasoning in terms of *fulfilment* to mend this problem. Simultaneously, our approach reduced the complexity present in the implication–based CRI–GMP without sacrifising any of its logical properties. As was exemplified in [7], it generally yields a fairly tight upper approximation of the corresponding CRI–GMP result and in a lot of cases, the results are equal.

We also generalized our method to cover a collection of parallel rules, as is typically the case in realistic applications. We reviewed some aggregation procedures and checked their suitability in the light of criteria such as coherency, consistency and speed. Options for future work include, amongst others, a more in–depth study of the notions of conflict and subnormalization, and how to deal with them effectively, as was done by Yager in [19].

## 5 Acknowledgements

Chris Cornelis would like to acknowledge the Fund for Scientific Research Flanders (FWO) for supporting the research elaborated on in this paper.

## References

1. Baldwin, J. F., Pilsworth, B. (1980): Axiomatic approach to implication for approximate reasoning using fuzzy logic. Fuzzy Sets and Systems **3**, 193–219
2. Bandler, W., Kohout, L. (1980): Fuzzy power sets and fuzzy implication operators. Fuzzy Sets and Systems **4**, 13–30
3. Bouchon–Meunier, B., Dubois, D., Godo, L., Prade, H. (1999): Fuzzy sets and possibility theory in approximate and plausible reasoning. Fuzzy sets in approximate reasoning and information systems, Kluwer Academic Publishers, 15–190
4. Bouchon–Meunier, B., Valverde, L. (1999): A fuzzy approach to analogical reasoning. Soft Computing **3**, 141–147

5. Buckley, J. J., Hayashi, Y. (1994): Can approximate reasoning be consistent? Fuzzy Sets and Systems **65**, 13–18
6. Cornelis, C., De Cock, M., Kerre, E. E. (2000): The generalized modus ponens in a fuzzy set theoretical framework. Fuzzy IF-THEN Rules in Computational Intelligence, Theory and Applications (eds. D. Ruan and E. E. Kerre), Kluwer Academic Publishers, 37–59
7. Cornelis, C., Kerre, E. E. (2001): Inclusion–based approximate reasoning. Lecture Notes in Computer Science 2074, (eds. V. Alexandrov, J. J. Dongarra, B. A. Juliano, R. S. Renner, C. J. K. Tan), Springer–Verlag, 200–210
8. Cornelis, C., Van Der Donck, C., Kerre, E. E. (2002): Sinha–Dougherty approach to the fuzzification of set inclusion revisited. Submitted to Fuzzy Sets and Systems
9. Dvořák, A. (2000): On preselection of rules in fuzzy logic deduction. International Journal of Uncertainty, Fuzziness and Knowledge–Based Systems **8(5)**, 563–573
10. Frago, N. (1996): Morfologia matematica borrosa basada en operadores generalizados de Łukasiewicz: procesiamento de imagines. Ph.D. thesis, Universidad publica de Navarra
11. Fukami, S., Mizumoto, M., Tanaka, T. (1981): Some considerations on fuzzy conditional inference. Fuzzy Sets and Systems **4**, 243-273
12. Klir, G. J., Yuan, B. (1995): Fuzzy sets and fuzzy logic, theory and applications. Prentice Hall
13. Klawonn, F., Novak, V. (1996): The relation between inference and interpolation in the framework of fuzzy systems. Fuzzy sets and systems **81**, 331–354
14. Ruan, D., Kerre, E. E. (1993): Fuzzy implication operators and generalized fuzzy method of cases. Fuzzy sets and systems **54**, 23–37
15. Sinha, D., Dougherty, E. R. (1993): Fuzzification of set inclusion: theory and applications. Fuzzy Sets and Systems **55**, 15–42
16. Türkşen, I. B., Zhong, Z. (1990): An approximate analogical reasoning scheme based on similarity measures and interval-valued fuzzy sets. Fuzzy Sets and Systems **34**, 323–346
17. Van Der Donck, C. (1998): A study of various fuzzy inclusions. Masters' thesis (in Dutch). Ghent University
18. Willmott, R. (1981): Mean measures of containment and equality between fuzzy sets. Proc. of the 11th Int. Symp. on Multiple–Valued Logic, Oklahoma City, 183–190
19. Yager, R. (2000): Approximate reasoning and conflict resolution. International Journal of Approximate Reasoning **25**, 15–42
20. Young, V. R. (1996): Fuzzy subsethood. Fuzzy Sets and Systems **77**, 371–384
21. Zadeh, L. A. (1971): Similarity relations and fuzzy orderings. Information Sciences **3**, 177–200
22. Zadeh, L. A. (1975): Calculus of fuzzy restrictions. Fuzzy sets and their applications to cognitive decision processes, Academic Press, 1–40