Fuzzy Control Revisited — Why Is It Working?

Joachim Weisbrod

IPD, Universität Karlsruhe
Vincenz–Priessnitz–Strasse 3
76133 Karlsruhe
Germany
weisbrod@ipd.info.uni-karlsruhe.de

Abstract

In the last decade fuzzy control (FC) has become a very popular control paradigm. Nowadays fuzzy controllers are very successful in a wide range of application domains and the market for such FC applications is still expanding. Nevertheless, the theoretical framework of FC is far beyond its practical success. While many convincing applications justify MAMDANI's original FC design, from a theoretical point of view there is still a simple question to be answered: Why is FC working? And to answer this question is not only a task to satisfy theorists but also a matter of practice. That is, to give FC a sound theoretical background means to answer a very important related question, namely: When is FC working? In which contexts may we expect FC to work properly and how should we prepare the controller inputs and interpret the given results?

In this paper we introduce a complete and sound framework to explain the most common FC mechanism, MAMDANI's original approach. We present the idea of σ–distributions, which is a concept somewhat dual to possibility theory, and prove that regarding fuzzy sets as σ–distributions means to postulate FC in the sense of MAMDANI. We point out that dealing with σ–distributions is based on dissonant bodies of evidence and therefore not the same as the use of possibility theory. Furthermore, with respect to incomplete or inconsistent knowledge both approaches show complementary characteristics, which is a very important observation from a practical point of view.

1 Introduction

Based on his pioneer paper [19] L.A. ZADEH initiated the still lasting era of fuzzy control (FC) in 1973 [20, 21]. The first application of fuzzy set theory to control is reported by MAMDANI and ASSILIAN in 1975 [10]. Their heuristic method turned out to be very successful, efficient, and easy to desribe, and therefore the vast majority of FC applications is using this method. For that reason we call it *standard fuzzy control* (SFC). But despite of its great practical success, SFC still lacks a sound theoretical background.

In this paper we introduce a complete and sound framework to explain SFC. We present the concept of σ–distributions and prove that regarding fuzzy sets as σ–distributions means to postulate SFC. This is, by giving the linguistic terms used to specify a fuzzy controller a concise but intuitive interpretation we verify the mechanisms applied by SFC. Furthermore, this new semantics may give the knowledge engineer trying to design a fuzzy controller a better understanding of his job and thereby it may help to accelerate and simplify FC design in general.

Our approach has been triggered by work of DUBOIS and PRADE. They have proposed exactly the same idea in [2, 4], but with different semantics. Interpreting fuzzy sets as possibility distributions [21, 22] they infer an inference mechanism for evidential reasoning. But this inference mechanism is very different from the one used in SFC. Nevertheless, it turns out that possibility distributions and σ–distributions are closely related to each other. A possibility distribution is equivalent to a corresponding fuzzy measure, namely a possibility measure and its dual necessity measure. A σ–distribution at least qualitatively corresponds to a more general fuzzy measure, namely a plausiblity measure and its dual belief measure. Therefore, with our new approach we have now two different inference mechanisms at hand to deal with qualitative knowledge.

With respect to incomplete or inconsistent knowledge both approaches show complementary characteristics: possibility distributions are very appropriate for dealing with incomplete knowledge, but show severe limitations when applied to inconsistent knowledge bases. For σ–distributions we find exactly the opposite properties. This is, both approaches complement each other perfectly.

This paper is organized as follows. Section 2 briefly describes SFC and lists some basic notation and terminology. In section 3 we recapitulate possibility theory and its effects on qualitative reasoning. The concept of σ–distributions is presented in section 4. In section 5 we infer the SFC inference mechanism from the concise new interpretation introduced. Section 6 deals with the relationship between those two ways to quantify qualitative knowledge and fuzzy measures. In section 7 we present some considerations concerning the basic characteristics of possibility and σ–distributions w.r.t inconsistent or incomplete knowledge bases. Finally, section 8 summarizes our work and gives some directions for future research.

2 Standard Fuzzy Control

A fuzzy controller is a rule based system. It consists of a set of rules, which are applied to the actual controller input to infer the controller output. This section recapitulates this inference mechanism and describes the special case of SFC. First of all, we give a few notations and definitions.

Without loss of generality we will consider only a *one input/one output* system. All results are very easily extended to systems with many input and many output variables and/or internal variables. Therefore, during this paper we will consider just two variables, an input variable x and an output variable y, with their *finite* universes of discourse being $\mathcal{U}_x$ and $\mathcal{U}_y$, respectively. Additionally, we denote the generic elements of $\mathcal{U}_x$ and $\mathcal{U}_y$ by u and v, respectively.

Definition 1 *A fuzzy set $\tilde{A}$ is called a* (fuzzy) subset *of universe $\mathcal{U}$ iff its membership function μ_A has domain $\mathcal{U}$: $\mu_A : \mathcal{U} \to [0,1]$. We denote this relation briefly by $\tilde{A} \subset \mathcal{U}$.*

During this paper, let $\tilde{A}$, $\tilde{A}'$, $\tilde{A}_i$ denote fuzzy subsets of $\mathcal{U}_x$, let $\tilde{B}$, $\tilde{B}'$, $\tilde{B}_i$, $\tilde{B}'_i$ denote fuzzy subsets of $\mathcal{U}_y$, and let $\tilde{R}$, $\tilde{R}_i$ denote fuzzy subsets of $\mathcal{U}_x \times \mathcal{U}_y$ $(i \in \mathbb{N})$.

Definition 2 *We call the linguistic expression* "variable x is fuzzy set $\tilde{A}$" *with $\tilde{A} \subset \mathcal{U}_x$, that sets variable x in relation to fuzzy set $\tilde{A}$, a* fuzzy predicate. *We denote this relation briefly by "x is $\tilde{A}$".*

The premise and consequent of each controller rule are built of such fuzzy predicates. Consequently, a controller rule reads

"IF x is $\tilde{A}$ THEN y is $\tilde{B}$",

or briefly $[\tilde{A} \Rightarrow \tilde{B}]$.

Now, the fundamental problem to be tackled by any FC mechanism is well known as *general modus ponens* (GMP). This is, given rule *"IF x is $\tilde{A}$ THEN y is $\tilde{B}$"* and an actual input *"x is $\tilde{A}'$"* [1], how are we do determine a conclusion *"y is $\tilde{B}'$"* inferred from applying $[\tilde{A} \Rightarrow \tilde{B}]$ to $\tilde{A}'$? The generally accepted answer is to use the *compositional rule of inference.* This is, to

[1] Note that a precise, crisp input value $u_0 \in \mathcal{U}_x$ may also be expressed as a fuzzy set.

represent rule $[\tilde{A} \Rightarrow \tilde{B}]$ as a *fuzzy implication relation* $\tilde{R} \subset \mathcal{U}_x \times \mathcal{U}_y$ and to compute $\tilde{B}'$ via

$$\tilde{B}' := \tilde{A}' \circ \tilde{R}, \tag{1}$$

where $\circ$ represents any s–t composition. For instance, using *max* and *min* as s– and t–norm, respectively, this leads to

$$\forall v \in \mathcal{U}_y : \mu_{B'}(v) := \max_{u \in \mathcal{U}_x} \{\min\{\mu_{A'}(u), \mu_R(u,v)\}\}.$$

In the next step, a method to handle the evaluation of n rules $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ has to be specified. The first method, called *FITA approach* ("first infer, then aggregate"), separately applies each single rule $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ to the input $\tilde{A}'$ to get n provisional results $\tilde{B}_i' = \tilde{A}' \circ \tilde{R}_i$. The derived fuzzy sets $\tilde{B}_i'$ are combined afterwards using an appropriate s– or t–norm. The so called *FATI approach* works just the other way round: First of all, using an appropriate s– or t–norm n relations $\tilde{R}_i$ representing the n rules $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ are aggregated. This aggregated relation $\tilde{R}$ is then used to compute $\tilde{B}' = \tilde{A}' \circ \tilde{R}$. In general, both approaches yield different results $\tilde{B}'$.

This universal FC skeleton obviously leaves a lot of open questions, like

- how to choose the implication relation $\tilde{R}$ given the rule $[\tilde{A} \Rightarrow \tilde{B}]$?
- which norms to apply in s–t composition?
- which approach to use in order to evaluate several rules? FITA or FATI?
- how to aggregate several implication relations $\tilde{R}_i$?
- how to aggregate several seperately computed results $\tilde{B}_i'$?

Combining every conceivable answer to any of these questions we get a vast amount of possible mechanisms. For further details we reference to [1, 9, 15], e.g.

On the other hand, it has already been mentioned, that there is an inference mechanism that is by far the most common one: SFC, the mechanism presented in 1975 by MAMDANI and ASSILIAN [10]. We are convinced that this approach originally was a purely heuristic one that has been developed without thinking in terms of implication relations and s–t composition. But nevertheless, SFC may be expressed in terms of the universal skeleton presented above:

In order to represent a given rule $[\tilde{A} \Rightarrow \tilde{B}]$ by a fuzzy implication relation $\tilde{R}$, SFC uses the minimum relation

$$\mu_R := \min(\mu_A, \mu_B), \tag{2}$$

which is just an abbreviated notation[2] for

$$\forall (u,v) \in \mathcal{U}_x \times \mathcal{U}_y : \mu_R(u,v) := \min \{\mu_A(u), \mu_B(v)\}.$$

To aggregate rules and inputs *max–min* composition is applied. In SFC a rule base of n rules $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ is evaluated by building the union of n separately evaluated conclusions $\tilde{B}_i'$ (FITA approach)[3]. All in all, we get

$$\tilde{B}' := \bigcup_i \tilde{B}_i' = \bigcup_i \left(\tilde{A}' \circ \tilde{R}_i\right), \tag{3}$$

where $\mu_{R_i} := \min(\mu_{A_i}, \mu_{B_i})$, $\circ$ represents *max–min* composition, and the union operator is realized by the *max*–norm.

Together with (2), equation (3) completely describes SFC. In the following we will show, that by giving the fuzzy predicates used to specify a fuzzy controller a new and concise interpretation all SFC design decisions above are not only justified but postulated. Within this paper we do not deal with defuzzification as we regard defuzzification as a feature that is just appended to the SFC inference mechanism. But the simple semantical model of σ–distributions will provide some useful hints on appropriate and inappropriate defuzzification mechanisms, as well.

3 Possibility Theory

It has already been mentioned that our argumentation finds its dual counterpart in possibility theory. There, assuming nothing but the simple semantics of possibility distributions a definite inference mechanism to deal with possibilistic knowledge may be inferred. Our approach leading from the semantics of σ–distributions directly to SFC uses exactly the same derivation steps. For that reason and in order to compare both approaches afterwards, we discuss this analog possibilistic derivation first, as presented concisely in [4] or in a more elaborated way in [14].

The basic idea is to regard the fuzzy predicates employed in the rules and inputs of a fuzzy system as possibility distributions. Setting $\pi_x := \mu_A$ we interpret the fuzzy predicate "*x is $\tilde{A}$*" as a flexible restriction (*elastic constraint*) π_x on the variable x:

[2]During this paper, we will strictly distinguish between $\min\{.\}$ and $\min(.)$. The further operation just selects the minimal element of its arguments, whereas the latter one constructs the pointwise minimum of its arguments. The same applies to maximization, naturally.

[3]Actually, within the SFC framework the FITA and FATI approach make no difference, since $\circ$ is distributive over $\cup$.

Definition 3 *A* possibility distribution (on $\mathcal{U}_x$) *is a mapping*

$$\pi_x : \mathcal{U}_x \to [0, 1],$$

that describes our imprecise knowledge on the real value of variable x by the following semantics:
$\pi_x(u)$ states the degree, to which we consider the assumption $x = u$ to be possible.

This is,

$$\pi_x(u) = 1 \quad \Leftrightarrow \quad x = u \text{ is (totally) possible,}$$
$$\pi_x(u) = 0 \quad \Leftrightarrow \quad x = u \text{ is impossible,}$$

with values $0 < \pi_x(u) < 1$ representing gradual evidence.

Furthermore, assuming reliable knowledge without contradictions there should be at least one $u_0 \in \mathcal{U}_x$, such that $\pi_x(u_0) = 1$. At least one value in $\mathcal{U}_x$ has to be possible without any doubt.

Definition 4 *A possibility distribution π_x is called* normalized *iff*

$$\exists u_0 \in \mathcal{U}_x : \pi_x(u_0) = 1.$$

The impact of the following definiton has first been stressed by YAGER [18]:

Definition 5 *A possibility distribution π'_x is at least as* specific *as possibility distribution π_x iff*

$$\forall u \in \mathcal{U}_x : \pi'_x(u) \leq \pi_x(u).$$

We denote this relation briefly by $\pi'_x \leq \pi_x$.

Whenever there are two valid pieces of information π'_x and π_x on the state of x with $\pi'_x \leq \pi_x$, π_x is redundant and may be dropped [4]. Specificity increases as the available information increases. We cite DUBOIS and PRADE from [4],p.147: "In that sense any possibility distribution π_x is provisional in nature and likely to be improved by further information, when the available one is not complete."

But how to combine two valid pieces of possibilistic information π_x^1 and π_x^2 in the general case, when $\pi_x^1 \not\leq \pi_x^2$ and $\pi_x^2 \not\leq \pi_x^1$? It is obvious, that two flexible restrictions π_x^1 and π_x^2 on x are expected to establish another flexible restriction π_x, that is to be at least as specific as both π_x^1 and π_x^2. That is, only distributions π_x obeying both $\pi_x \leq \pi_x^1$ and $\pi_x \leq \pi_x^2$ are compatible with both π_x^1 and π_x^2. Using the intuitive *principle of minimum specificity*[4] we choose the maximal distribution compatible with both π_x^1 and π_x^2 and find

$$\begin{aligned} \pi_x &:= \max\{\pi'_x \mid \pi'_x \leq \pi_x^1 \ \wedge \ \pi'_x \leq \pi_x^2\} \\ &= \min(\pi_x^1, \pi_x^2). \end{aligned} \tag{4}$$

This is, the resulting distribution π_x is to regard any constraint contained within π_x^1 and π_x^2, but must not make any additional assumptions. Any distribution π'_x with $\pi'_x < \pi_x$ is more precise than allowed by the knowledge π_x^1 and π_x^2.

With (joint) possibility distributions

$$\pi_{x_1,\ldots,x_n} : \mathcal{U}_{x_1} \times \cdots \times \mathcal{U}_{x_n} \to [0,1]$$

we can express vague knowledge about relations between several variables x_1 to x_n, as well: For instance, $\pi_{x,y}(u,v)$ states the degree, to which we consider the assumption $(x,y) = (u,v)$, i.e. both $x = u$ and $y = v$, to be possible (cf. definition 3, with $\mathcal{U} := \mathcal{U}_x \times \mathcal{U}_y$). This is, possibility distributions may be specified on arbitrary joint universes without any semantical complications.

Given a joint possibility distribution, say $\pi_{x,y}$, we may also be interested in the meaning of this distribution on the separated variables x or y, respectively.

Definition 6 *The* projection $[\pi_{x,y} \downarrow \mathcal{U}_x]$ *of the possibility distribution* $\pi_{x,y} : \mathcal{U}_x \times \mathcal{U}_y \to [0,1]$ *on* $\mathcal{U}_x$ *is defined by*

$$\begin{aligned} [\pi_{x,y} \downarrow \mathcal{U}_x] \ : \ \mathcal{U}_x &\to [0,1] \\ u &\mapsto \max_{v \in \mathcal{U}_y}\{\pi_{x,y}(u,v)\}. \end{aligned}$$

This is, in order to calculate the projection $[\pi_{x,y} \downarrow \mathcal{U}_x]$ of the joint distribution $\pi_{x,y}$ on $\mathcal{U}_x$ we assume to know nothing about the real value of y. Therefore we assign to $u \in \mathcal{U}_x$ the highest possibility grade $\pi_{x,y}(u,v)$ that is allowed by any $v \in \mathcal{U}_y$. And since we assume that there exists a $v' \in \mathcal{U}_y$ such that $y = v'$, the flexible restriction $[\pi_{x,y} \downarrow \mathcal{U}_x]$ derived for variable x is correct and cannot be too specific.

Proceeding from these considerations it is very straightforward and natural to derive the so called *combination/projection principle*, which is the

[4]Quoted from [4], as well: "More generally, when the available information stems from several sources that can be considered as reliable, the possibility distribution that accounts for it is the least specific possibility distribution that satisfies the set of constraints induced by the pieces of information given by the different sources."

universal mechanism to combine possibilistic information on arbitrary universes of discourse[5]. And this mechanism turns out to be nothing but *max–min* composition [14, 4].

At this point we come back to our rule based system. Dealing with possibilistic information, a rule *"IF x is $\tilde{A}$ THEN y is $\tilde{B}$"* corresponds to a possibility distribution $\pi^{\Rightarrow}_{x,y}$ defined on $\mathcal{U}_x \times \mathcal{U}_y$. An actual input *"$x$ is $\tilde{A}'$"* corresponds to another possibility distribution π'_x defined on $\mathcal{U}_x$. But within the possibilistic framework there is just one way to combine $\pi^{\Rightarrow}_{x,y}$ and π'_x in order to get an output π'_y: the derived combination/projection principle, namely

$$\pi'_y := \pi'_x \circ \pi^{\Rightarrow}_{x,y} \tag{5}$$

or in detail

$$\forall v \in \mathcal{U}_y : \pi'_y(v) := \max_{u \in \mathcal{U}_x} \{\min\{\pi'_x(u), \pi^{\Rightarrow}_{x,y}(u,v)\}\}.$$

And therefore, this is the *one and only* way to resolve the generalized modus ponens when interpreting fuzzy predicates as possibility distributions. But there is one last problem to be solved: How should we choose the joint distribution $\pi^{\Rightarrow}_{x,y}$ to represent rule $[\tilde{A} \Rightarrow \tilde{B}]$?

In order to derive $\pi^{\Rightarrow}_{x,y}$ we use just the most fundamental semantic property of a rule: If we apply rule $[\tilde{A} \Rightarrow \tilde{B}]$ to the input $\tilde{A}$ that exactly matches the rule premise, we obviously expect the rule consequent $\tilde{B}$ to be inferred. In terms of possibility distributions, this is:

$$\pi_x \circ \pi^{\Rightarrow}_{x,y} \stackrel{!}{=} \pi_y, \tag{6}$$

where $\pi_x := \mu_A$ and $\pi_y := \mu_B$; applying the exact premise π_x to $\pi^{\Rightarrow}_{x,y}$ is to infer exactly π_y. So from this point of view the set

$$\Omega := \{\pi_{x,y} \mid \pi_x \circ \pi_{x,y} = \pi_y\} \tag{7}$$

contains all possibility distributions $\pi_{x,y}$ able to represent the given rule. And the principle of minimum specificity tells us, which element to take from Ω: its maximal solution $\pi^*_{x,y}$, so it exists. Without any further assumptions we are not allowed to use any more specific $\pi'_{x,y} \in \Omega$ with $\pi'_{x,y} < \pi^*_{x,y}$.

It is easy to show, that whenever $\Omega \neq \emptyset$, $\pi^*_{x,y}$ is in fact unequivocally determined. In this case, we set

$$\pi^{\Rightarrow}_{x,y} := \pi^*_{x,y} := \max\{\pi_{x,y} \mid \pi_{x,y} \in \Omega\}. \tag{8}$$

[5] Actually, this principle is a generalization of ZADEH's well known *extension principle*.

It turns out, that $\pi^*_{x,y}$ is the well known GÖDEL relation specified by the α operator [12, 17]:

$$\pi^*_{x,y} \quad := \quad \pi_x \ \alpha \ \pi_y \tag{9}$$

$$:\Leftrightarrow \ \pi^*_{x,y}(u,v) \quad := \quad \begin{cases} 1 & \text{if } \pi_x(u) \leq \pi_y(v) \\ \pi_y(v) & \text{else} \end{cases} \tag{10}$$

Furthermore, it holds

$$\Omega \neq \emptyset \quad \Leftrightarrow \quad \max_{u \in \mathcal{U}_x} \{\pi_x(u)\} \geq \max_{v \in \mathcal{U}_y} \{\pi_y(v)\}. \tag{11}$$

Therefore, whenever π_x and π_y are normalized, which is a very natural requirement, the GÖDEL relation $\pi_x \ \alpha \ \pi_y$ is the right choice to represent our rule under the possibilistic interpretation.

To apply a set of n rules $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ now becomes a very easy task. First, for any $i \leq n$ we set $\pi^i_x := \mu_{A_i}$, $\pi^i_y := \mu_{B_i}$ and determine the possibility distribution

$$\pi^{\Rightarrow i}_{x,y} \quad = \quad \max\{\pi_{x,y} \mid \pi_{x,y} \in \Omega_i\} \tag{12}$$

$$\quad = \quad \max\{\pi_{x,y} \mid \pi^i_x \circ \pi_{x,y} = \pi^i_y\} \tag{13}$$

to represent rule i by calculating the corresponding GÖDEL relation $\pi^i_x \ \alpha \ \pi^i_y$.

And with n rules $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ being nothing but n possibilistic constraints $\pi^{\Rightarrow i}_{x,y}$ on the relation between x and y it is obvious how to handle such a rule base: Applying the principle of minimum specificity (4), in the second step we build a *meta rule*

$$\pi^{\Rightarrow}_{x,y} := \min_i \, (\pi^{\Rightarrow i}_{x,y}) \tag{14}$$

that considers all the information given by the rule base.

Finally, in order to apply this meta rule $\pi^{\Rightarrow}_{x,y}$ to a given actual input information π'_x we obtain a conclusion π'_y by using the combination/projection principle (1):

$$\pi'_y := \pi'_x \circ \pi^{\Rightarrow}_{x,y}. \tag{15}$$

With this mechanism we are able to handle arbitrary rule bases, given that we regard the fuzzy predicates involved as possibility distributions. We will refer to this very well substantiated, non heuristic approach as π*–reasoning*.

4 σ–Distributions

In the preceeding section we showed, that an unequivocal mechanism to tackle the generalized modus ponens can be inferred from just assuming the very simple and intuitive semantics of possibility distributions. We now introduce the new concept of *σ–distributions* in order to infer SFC using a derivation very similar to the one presented above.

A σ–distribution $\sigma_x : \mathcal{U}_x \to [0,1]$ specifies our imprecise knowledge on the state of variable x from a different point of view than a possibility distribution, i.e. under a different interpretation:

Definition 7 *A* σ–distribution (on $\mathcal{U}_x$) *is a mapping*

$$\sigma_x : \mathcal{U}_x \to [0,1],$$

that describes our imprecise knowledge on the real value of variable x by the following semantics:
$\sigma_x(u)$ states the degree, to which the assumption $x = u$ is supported *by the available knowledge.*

This is,

$$\begin{aligned} \sigma_x(u) = 1 \quad &\Leftrightarrow \quad x = u \text{ is (totally) supported,} \\ \sigma_x(u) = 0 \quad &\Leftrightarrow \quad x = u \text{ is not supported at all,} \end{aligned}$$

with values $0 < \sigma_x(u) < 1$ representing gradual evidence.

Corresponding to the notion of specificity in the possibilistic case, we define completeness:

Definition 8 *A σ–distribution σ'_x is at least as* complete *as σ–distribution σ_x iff*

$$\forall u \in \mathcal{U}_x : \sigma'_x(u) \geq \sigma_x(u).$$

We denote this relation briefly by $\sigma'_x \geq \sigma_x$.

In this case, combining two pieces of information σ_x^1 and σ_x^2 leads to a σ–distribution σ_x, that is to be at least as complete as both σ_x^1 and σ_x^2, i.e. $\sigma_x \geq \sigma_x^1$ and $\sigma_x \geq \sigma_x^2$. Taking into account the obvious requirement, that the resulting σ_x is to carry no information contained neither in σ_x^1 nor in σ_x^2, we postulate an analogue to the principle of minimum specificity. We call it *principle of minimum completeness*. This principle directs us to select the minimal element from the set of σ–distributions compatible to the present knowledge. So we derive dual to (4)

$$\begin{aligned} \sigma_x \;&:=\; \min\{\sigma'_x \mid \sigma'_x \geq \sigma_x^1 \;\wedge\; \sigma'_x \geq \sigma_x^2\} \\ &=\; \max(\sigma_x^1, \sigma_x^2). \end{aligned} \tag{16}$$

We have to be careful with joint σ–distributions defined on joint universes. Following definition 7, with $\sigma_{x,y}$ being a joint σ–distribution

$$\sigma_{x,y} : \mathcal{U}_x \times \mathcal{U}_y \to [0,1],$$

$\sigma_{x,y}(u,v)$ states the degree, to which the assumption $(x,y) = (u,v)$ is supported by the available knowledge. $\sigma_{x,y}(u,v)$ quantifies the degree of support, that at the same time it holds $x = u$ and $y = v$. But in general, the joint distribution $\sigma_{x,y}$ has no meaning on the current states of the separate variables x and y, respectively. For instance, high support for the assumption, that at the same time the weather may be hot and the beach may be crowded has no effect on the current weather at all. Speaking in terms of possibility theory, the *projection* of $\sigma_{x,y}$ on either $\mathcal{U}_x$ or $\mathcal{U}_y$ is a distribution $\sigma \equiv 0$, in general. We define:

Definition 9 *The* projection $[\sigma_{x,y} \downarrow \mathcal{U}_x]$ *of the* σ*–distribution* $\sigma_{x,y} : \mathcal{U}_x \times \mathcal{U}_y \to [0,1]$ *on* $\mathcal{U}_x$ *is defined by*

$$\begin{array}{rcrcl} [\sigma_{x,y} \downarrow \mathcal{U}_x] & : & \mathcal{U}_x & \to & [0,1] \\ & & u & \mapsto & \min\limits_{v \in \mathcal{U}_y} \{\pi_{x,y}(u,v)\}. \end{array}$$

This is, in order to get the degree of support $[\sigma_{x,y} \downarrow \mathcal{U}_x](u)$ for the assumption $x = u$ we have to be pessimistic: since $\sigma_{x,y}$ tells us nothing about the state of the variable y, y may take any value $v \in \mathcal{U}_y$. So we have to look for the element $v' \in \mathcal{U}_y$ with the lowest degree of support $\sigma_{x,y}(u,v')$ for the assumption $(x,y) = (u,v')$. This degree is the best we can get to support the assertion $x = u$.

But what happens, if there is in fact some further information on x or y, respectively? In this case we have to deal with two σ–distributions: a joint distribution $\sigma_{x,y}$ and additional information on x given by distribution σ_x, for instance. What can be inferred about the variable y in this case? If there are values $u \in \mathcal{U}_x$ and $v \in \mathcal{U}_y$, such that $\sigma_x(u) > 0$ and $\sigma_{x,y}(u,v) > 0$, there is obviously some support for the assumption $y = v$: $\sigma_x(u)$ supports the assumption $x = u$, whereas $\sigma_{x,y}(u,v)$ supports the assumption, that at the same time both $x = u$ and $y = v$ hold. Since the degree τ_u of the support for $y = v$ depends on the interaction between $\sigma_x(u)$ and $\sigma_{x,y}(u,v)$, it must fulfill either $\tau_u \geq \sigma_x(u)$ or $\tau_u \geq \sigma_{x,y}(u,v)$. Applying the principle of minimum completeness, we look for the smallest value to satisfy this condition and find

$$\begin{array}{rcl} \tau_u : & = & \min\{\tau' \mid \tau' \geq \sigma_x(u) \ \vee \ \tau' \geq \sigma_{x,y}(u,v)\} \\ & = & \min\{\sigma_x(u), \sigma_{x,y}(u,v)\}. \end{array} \tag{17}$$

To get the overall support $\sigma_y(v)$ for the assumption $y = v$ we have to regard all the elements of $\mathcal{U}_x$ and using the principle of minimum completeness (16) once again, this is

$$\begin{aligned}\sigma_y(v) &= \max_{u \in \mathcal{U}_x} \{\tau_u\} \\ &= \max_{u \in \mathcal{U}_x} \{\min\{\sigma_x(u), \sigma_{x,y}(u,v)\}\}.\end{aligned} \tag{18}$$

And to get all the information derivable for y from σ_x and $\sigma_{x,y}$ we build σ_y by calculating (18) for all $v \in \mathcal{U}_y$. But this is nothing but to evaluate

$$\sigma_y = \sigma_x \circ \sigma_{x,y}, \tag{19}$$

with $\circ$ representing ordinary *max–min* composition.

So again we have found *max–min* composition. And again, this turns out to be the universal mechanism to combine information in terms of σ–distributions on arbitrary universes.

5 From σ–Distributions to SFC

So far we have introduced the semantics of σ–distributions and the universal way how to combine them. Now, we come back to the inference mechanism of a rule based system, this time by means of σ–distributions.

Assuming $\sigma_x := \mu_A$, a σ–distribution

$$\sigma_x : \mathcal{U}_x \to [0,1]$$

is just another way to interpret the fuzzy predicate *"x is $\tilde{A}$"*. But how to represent a rule $[\tilde{A} \Rightarrow \tilde{B}]$ in this context? As in the possibilistic case we have to find a joint distribution $\sigma_{x,y}$ appropriate to represent the meaning of a rule, but this time according to the semantics of σ–distributions.

We regard $[\tilde{A} \Rightarrow \tilde{B}]$ as a piece of supporting information

$$\sigma^{\Rightarrow}_{x,y} : \mathcal{U}_x \times \mathcal{U}_y \to [0,1]$$

on the relation between x and y. Again, we infer the distribution $\sigma^{\Rightarrow}_{x,y}$ by considering the most fundamental semantics of the linguistic implication *"IF x is $\tilde{A}$ THEN y is $\tilde{B}$"*: Applying $[\tilde{A} \Rightarrow \tilde{B}]$ to the input $\tilde{A}$ is to infer exactly $\tilde{B}$. In terms of the corresponding σ–distributions, this is:

$$\sigma_x \circ \sigma^{\Rightarrow}_{x,y} \stackrel{!}{=} \sigma_y, \tag{20}$$

where $\sigma_x := \mu_A$ and $\sigma_y := \mu_B$. So once more, we examine the set

$$\Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}. \tag{21}$$

But obeying the principle of minimum completeness, this time we have to look for a *minimal* element in Ω, instead of the maximal one chosen in the possibilistic case. It is well known, that in general Ω embodies a set $\Omega_* \subset \Omega$ of *several* minimal elements [8]. With $\sigma'_{x,y} \in \Omega_*$ and $\sigma''_{x,y} \in \Omega_*$ it holds $\sigma'_{x,y} \not\leq \sigma''_{x,y}$ and $\sigma''_{x,y} \not\leq \sigma'_{x,y}$. Lemma 10 characterizes the elements of Ω_* without explicitely determining them:

Lemma 10 *Let* $\Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}$ *and* $\Omega_* \subset \Omega$ *be the set of minimal elements of* Ω. *Then it is:*

$$\begin{array}{llrl} \sigma_{x,y} \in \Omega_* & \textit{iff} & \forall v \in \mathcal{U}_y : & \exists u' \in \mathcal{U}_x : \\ & & & \sigma_{x,y}(u', v) = \sigma_y(v) \\ & & \wedge & \sigma_x(u') \geq \sigma_y(v) \\ & & \wedge & [u \neq u' \Rightarrow \sigma_{x,y}(u, v) = 0]. \end{array}$$

Proof: "$\Rightarrow$": with $\sigma_{x,y} \in \Omega_* \subset \Omega$ it is $\sigma_x \circ \sigma_{x,y} = \sigma_y$. Therefore for any $v \in \mathcal{U}_y$ it holds:

$$\max_{u \in \mathcal{U}_x} \{\min\{\sigma_x(u), \sigma_{x,y}(u, v)\}\} = \sigma_y(v)$$

$$\begin{array}{ll} \Rightarrow & \exists u' \in \mathcal{U}_x : \min\{\sigma_x(u'), \sigma_{x,y}(u', v)\} = \sigma_y(v) \\ & \wedge[u \neq u' \Rightarrow \min\{\sigma_x(u), \sigma_{x,y}(u, v)\} \leq \sigma_y(v)] \\ \Rightarrow & \exists u' \in \mathcal{U}_x : \sigma_{x,y}(u', v) \geq \sigma_y(v) \ \wedge \ \sigma_x(u') \geq \sigma_y(v) \\ & \wedge[u \neq u' \Rightarrow \min\{\sigma_x(u), \sigma_{x,y}(u, v)\} \leq \sigma_y(v)] \end{array}$$

Together with the requirement that $\sigma_{x,y}$ is to be minimal in Ω the claim is verified.

"$\Leftarrow$": for all $v \in \mathcal{U}_y$ the existence of a distribution $\sigma_{x,y}$ with $\exists u' \in \mathcal{U}_x$ s.t. $\sigma_{x,y}(u', v) = \sigma_y(v)$, $\sigma_x(u') \geq \sigma_y(v)$ and $[u \neq u' \Rightarrow \sigma_{x,y}(u, v) = 0]$ implies

$$\begin{array}{ll} \multicolumn{2}{l}{\max_{u \in \mathcal{U}_x} \{\min\{\sigma_x(u), \sigma_{x,y}(u, v)\}\}} \\ = & \min\{\sigma_x(u'), \sigma_{x,y}(u', v)\} \\ = & \sigma_y(v). \end{array}$$

Therefore it is $\sigma_{x,y} \in \Omega$. Now we assume $\sigma_{x,y} \notin \Omega_*$:

$$\begin{array}{lll} \Rightarrow & & \exists \sigma'_{x,y} \in \Omega : \sigma'_{x,y} < \sigma_{x,y} \\ \Rightarrow & & \exists (u', v') \in \mathcal{U}_x \times \mathcal{U}_y : \sigma'_{x,y}(u', v') < \sigma_{x,y}(u', v') \\ & \wedge & \forall (u, v) \in \mathcal{U}_x \times \mathcal{U}_y : \sigma'_{x,y}(u, v) \leq \sigma_{x,y}(u, v) \\ \Rightarrow & & \sigma_{x,y}(u', v') = \sigma_y(v') \ \wedge \ \sigma'_{x,y}(u', v') < \sigma_y(v') \\ & \wedge & [u \neq u' \Rightarrow \sigma_{x,y}(u, v') = \sigma'_{x,y}(u, v') = 0] \\ \Rightarrow & & \max_{u \in \mathcal{U}_x} \{\min\{\sigma_x(u), \sigma'_{x,y}(u, v')\}\} < \sigma_y(v') \end{array}$$

This is a contradiction to $\sigma'_{x,y} \in \Omega$ and therewith the assumption $\sigma_{x,y} \notin \Omega_*$ must be wrong. ■

In order to find $\sigma^{\Rightarrow}_{x,y}$, at first glance we are tempted to choose one distinct element $\sigma'_{x,y}$ from Ω_*. But lemma 10 shows, that applying the chosen distribution $\sigma'_{x,y}$ to σ_x by computing $\sigma_x \circ \sigma'_{x,y}$, for each $v \in \mathcal{U}_y$ there is just one $u' \in \mathcal{U}_x$ that is responsible for the support $\sigma_y(v)$ derived for the assumption $y = v$. All the other elements $u \neq u'$ in $\mathcal{U}_x$ do not contribute anything to support the assumption $y = v$, since $\sigma'_{x,y}(u, v) = 0$. If this is, what the designer of the rule base had in mind, there would have been no need for designing a *fuzzy* rule base. And even if this was intended, he did not specify which one of the minimal solutions to choose. So we must suppose that he meant all of them. Therefore each element of Ω_* is a piece of information supported by the semantics of the given rule.

And by now we know how to combine several valid σ–distributions. Using the principle of minimum completeness (16) we build the union

$$\sigma_{x,y*} := \max_{\sigma_{x,y} \in \Omega_*} (\sigma_{x,y}). \tag{22}$$

of all the minimal solutions.

Definition 11 *Let $\Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}$ and let $\Omega_* \subset \Omega$ denote the set of minimal elements from Ω. Then we call the σ–distribution*

$$\sigma_{x,y*} := \max_{\sigma_{x,y} \in \Omega_*} (\sigma_{x,y})$$

the minimal general solution *of Ω.*

But does this minimal general solution $\sigma_{x,y*}$ still fulfill our fundamental requirement $\sigma_x \circ \sigma_{x,y*} = \sigma_y$? To prove that $\sigma_{x,y*} \in \Omega$ we present the following lemma:

Lemma 12 *Let $\sigma_{x,y} \in \Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}$ and let $\sigma^*_{x,y}$ be the maximal element of Ω. Then it holds:*

$$\forall \sigma'_{x,y} : \sigma_{x,y} \leq \sigma'_{x,y} \leq \sigma^*_{x,y} \;\Rightarrow\; \sigma'_{x,y} \in \Omega.$$

Proof: Let $\sigma_{x,y} \in \Omega$. Using the monotonicity of *max–min* composition, $\sigma_{x,y} \leq \sigma'_{x,y} \leq \sigma^*_{x,y}$ implies

$$\sigma_y = \sigma_x \circ \sigma_{x,y} \leq \sigma_x \circ \sigma'_{x,y} \leq \sigma_x \circ \sigma^*_{x,y} = \sigma_y.$$

Therefore it is $\sigma_x \circ \sigma'_{x,y} = \sigma_y$. ■

Now, since for any $\sigma'_{x,y} \in \Omega_*$ it is obvious that

$$\sigma'_{x,y} \leq \max_{\sigma_{x,y} \in \Omega_*} (\sigma_{x,y}) = \sigma_{x,y*} \leq \sigma^*_{x,y},$$

lemma 12 guarantees that $\sigma_{x,y*} \in \Omega$.

But in general, the set of minimal solutions Ω_* is very big and therefore it is very hard to compute the minimal general solution $\sigma_{x,y*}$ of Ω by means of definition 11 and equation (22). The following theorem points to a much easier way to obtain $\sigma_{x,y*}$ by means of the ϵ operator [12]:

Theorem 13 *Let $\Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}$. Then for the minimal general solution $\sigma_{x,y*}$ of Ω it holds*

$$\begin{aligned} \sigma_{x,y*} &:= \sigma_x \,\epsilon\, \sigma_y \\ :\Leftrightarrow \quad \sigma_{x,y*}(u,v) &:= \begin{cases} 0 & \text{if } \sigma_x(u) < \sigma_y(v) \\ \sigma_y(v) & \text{else.} \end{cases} \end{aligned}$$

Proof: (1) We show that $\sigma_{x,y*} \leq \sigma_x \,\epsilon\, \sigma_y$:
With lemma 10 for any $\sigma_{x,y} \in \Omega_*$ and $\forall (u,v) \in \mathcal{U}_x \times \mathcal{U}_y$ it holds:

$$\begin{aligned} \sigma_x(u) \geq \sigma_y(v) \quad &\Rightarrow \quad \sigma_{x,y}(u,v) \in \{0, \sigma_y(v)\} \\ &\Rightarrow \quad \sigma_{x,y}(u,v) \leq (\sigma_x \,\epsilon\, \sigma_y)(u,v) \\ \sigma_x(u) < \sigma_y(v) \quad &\Rightarrow \quad \sigma_{x,y}(u,v) = 0 \\ &\Rightarrow \quad \sigma_{x,y}(u,v) = (\sigma_x \,\epsilon\, \sigma_y)(u,v) = 0 \end{aligned}$$

But with $\forall \sigma_{x,y} \in \Omega_* : \sigma_{x,y} \leq \sigma_x \,\epsilon\, \sigma_y$ it is

$$\sigma_{x,y*} = \max_{\sigma_{x,y} \in \Omega_*} (\sigma_{x,y}) \leq \sigma_x \,\epsilon\, \sigma_y.$$

(2) We show that $\sigma_{x,y*} \geq \sigma_x \,\epsilon\, \sigma_y$:
Assumption: $\sigma_{x,y*} \not\geq \sigma_x \,\epsilon\, \sigma_y$

$$\begin{aligned} \Rightarrow \quad \exists (u',v') : \quad & \sigma_{x,y*}(u',v') < (\sigma_x \,\epsilon\, \sigma_y)(u',v') \\ \Rightarrow \quad \forall \sigma_{x,y} \in \Omega_* : \quad & \sigma_{x,y}(u',v') < (\sigma_x \,\epsilon\, \sigma_y)(u',v') \end{aligned}$$

Case 1. $\sigma_x(u') < \sigma_y(v')$:

$$\Rightarrow \forall \sigma_{x,y} \in \Omega_* : \; \sigma_{x,y}(u',v') < 0$$

This inequality obviously contradicts definition 7.

Case 2. $\sigma_x(u') \geq \sigma_y(v')$:

$$\Rightarrow \forall \sigma_{x,y} \in \Omega_* : \; \sigma_{x,y}(u',v') < \sigma_y(v')$$

With $\sigma_y(v') = 0$ there is already a contradiction, otherwise we can infer from lemma 10 the existence of another solution $\sigma'_{x,y} \notin \Omega_*$, for which it holds: $\sigma'_{x,y}(u', v') = \sigma_y(v')$ and $\forall u'' \neq u' : \sigma'_{x,y}(u'', v') = 0$. Now $\sigma'_{x,y}$ is not comparable to any other element from Ω_* and this is a contradiction to the precondition, that Ω_* contains all the minimal solutions from Ω. ■

But there is one more very basic semantic property inherent in a linguistic rule $[\tilde{A} \Rightarrow \tilde{B}]$. If input $\tilde{A}$ implies output $\tilde{B}$, then even more so any $\tilde{B}'$ is implied that is a generalization of $\tilde{B}$, i.e. any $\tilde{B}'$ that may be inferred from assuming nothing but $\tilde{B}$. And in terms of σ–distributions, this is any $\tilde{B}'$ that is less complete than $\tilde{B}$, i.e. with $\sigma_y := \mu_B$ any σ'_y with $\sigma'_y \leq \sigma_y$. Any σ'_y with $\sigma'_y \leq \sigma_y$ is a redundant piece of information that is implicitely contained within σ_y. Trivially, we may infer any information that is less complete than σ_y, as well.

This second fundamental semantic feature of a linguistic rule is considered in the possibilistic case, as well. There, with π_y we may also infer any possibility distribution $\pi'_y \geq \pi_y$ that is less specific. But since this additional property does not influence the derived result at all, the dual step was omitted in the possibilistic derivation presented in section 3 for the sake of clarity.

In order to consider this second basic rule property we have to examine all the sets

$$\Omega(\sigma'_y) := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma'_y\} \text{ with } \sigma'_y \leq \sigma_y. \tag{23}$$

For each of these sets $\Omega(\sigma'_y)$ we have to determine its minimal general solution

$$\sigma_{x,y*}(\sigma'_y) := \sigma_x \,\epsilon\, \sigma'_y. \tag{24}$$

And again, for all $\sigma'_y \leq \sigma_y$, the distribution $\sigma_{x,y*}(\sigma'_y)$ represents a piece of information supported by the given linguistic rule $[\tilde{A} \Rightarrow \tilde{B}]$. Using equation (16) once more in order to combine all the resulting distributions we will derive the minimum relation, finally. At first, we define:

Definition 14 *Let $\Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}$ and let $\sigma_{x,y*}(\sigma'_y)$ denote the minimal general solution of the set $\Omega(\sigma'_y) := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma'_y\}$. Then we call the σ–distribution*

$$\sigma^{\Rightarrow}_{x,y} := \max_{\sigma'_y : \sigma'_y \leq \sigma_y} (\sigma_{x,y*}(\sigma'_y))$$

the minimal representative solution *of the set Ω.*

The minimal representative solution $\sigma^{\Rightarrow}_{x,y}$ of Ω turns out to be the well known minimum relation $\min(\sigma_x, \sigma_y)$ used in SFC that has already been presented in equation (2):

Theorem 15 *Let* $\Omega := \{\sigma_{x,y} \mid \sigma_x \circ \sigma_{x,y} = \sigma_y\}$. *Then for the minimal representative solution* $\sigma_{x,y}^{\Rightarrow}$ *of* Ω *it holds:*

$$\sigma_{x,y}^{\Rightarrow} := \min(\sigma_x, \sigma_y).$$

Proof:

$$\begin{aligned}\sigma_{x,y}^{\Rightarrow} &\overset{\text{def. 14}}{=} \max_{\sigma'_y \leq \sigma_y} (\sigma_{x,y*}(\sigma'_y)) \\ &\overset{\text{theor. 13}}{=} \max_{\sigma'_y \leq \sigma_y} (\sigma_x \,\epsilon\, \sigma'_y)\end{aligned}$$

Therefore, for all $(u, v) \in \mathcal{U}_x \times \mathcal{U}_y$ it is:

$$\begin{aligned}\sigma_{x,y}^{\Rightarrow}(u,v) &= \left(\max_{\sigma'_y \leq \sigma_y} (\sigma_x \,\epsilon\, \sigma_y)\right)(u,v) \\ &= \max_{\sigma'_y \leq \sigma_y} \begin{cases} 0 & \text{falls } \sigma_x(u) < \sigma'_y(v) \\ \sigma'_y(v) & \text{falls } \sigma_x(u) \geq \sigma'_y(v) \end{cases}\end{aligned}$$

Case 1. $\sigma_x(u) \geq \sigma_y(v)$:

$$\Rightarrow \sigma_{x,y}^{\Rightarrow}(u,v) = \max_{\sigma'_y \leq \sigma_y} \sigma'_y(v) = \sigma_y(v)$$

Case 2. $\sigma_x(u) < \sigma_y(v)$:

$$\begin{aligned}&\Rightarrow \exists \sigma'_y < \sigma_y : \sigma_x(u) \geq \sigma'_y(v) \\ &\Rightarrow \sigma_{x,y}^{\Rightarrow}(u,v) = \max_{\sigma'_y : \sigma'_y < \sigma_y \wedge \sigma'_y(v) \leq \sigma_x(u)} \{\sigma'_y(v)\} = \sigma_x(u)\end{aligned}$$

Considering both case 1 and case 2 completes the proof. ∎

As already suggested by its notation, the minimal representative solution $\sigma_{x,y}^{\Rightarrow}$ of Ω is the appropriate σ–distribution to represent our rule. It is the least complete element of Ω that covers the basic semantics of the rule.

In order to apply n rules $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ we set $\sigma_x^i := \mu_{A_i}$ and $\sigma_y^i := \mu_{B_i}$ and represent any rule by the minimal representative solution $\sigma_{x,y}^{\Rightarrow i}$ of the corresponding solution set $\Omega_i := \{\sigma_{x,y} \mid \sigma_x^i \circ \sigma_{x,y} = \sigma_y^i\}$:

$$\sigma_{x,y}^{\Rightarrow i} := \min(\sigma_x^i, \sigma_y^i). \tag{25}$$

Since *max–min* composition is distributive over maximization it makes no difference, whether we use the FITA or the FATI approach (see section 2) in order to evaluate all rules. This is, applying n rules $\sigma_{x,y}^{\Rightarrow i}$ to input σ'_x we may evaluate each rule separately first and then combine the resulting valid pieces of information $\sigma_x \circ \sigma_{x,y}^{\Rightarrow i}$ by means of (16) or we combine the rules first by assigning

$$\sigma_{x,y}^{\Rightarrow} := \max_i \left(\sigma_{x,y}^{\Rightarrow i}\right) \tag{26}$$

and then evaluate the resulting meta rule $\sigma_{x,y}^{\Rightarrow}$ using *max–min* composition. It holds

$$\begin{aligned} \sigma'_y &= \sigma'_x \circ \max_i \left(\sigma_{x,y}^{\Rightarrow i}\right) && \text{(FATI)} \\ &= \max_i \left(\sigma'_x \circ \sigma_{x,y}^{\Rightarrow i}\right) && \text{(FITA)} \end{aligned} \tag{27}$$

Obviously, the FITA approach above matches exactly equation (3) representing SFC. This is, assuming nothing but the simple semantics of σ–distributions and the most fundamental semantic properties of a linguistic rule we were able to derive the very SFC inference mechanism. In the following we will refer to this approach as σ–reasoning instead of the term SFC, since a priori reasoning with σ–distributions is not restricted to control tasks.

6 σ–Distributions and Belief Measures

In order to get some more useful insight into π-reasoning, σ–reasoning and the link between them, it is very helpful to study their relation to fuzzy measure theory. To check the details that are very briefly presented in this section we refer to [13], [8], and [16].

Again, we start by presenting some well known facts about possibility distributions. Whenever π_x is *normalized*, there is a bijective mapping between possibility distributions $\pi_x : \mathcal{U}_x \to [0,1]$ and *possibility measures* $\Pi_x : 2^{\mathcal{U}_x} \to [0,1]$ by means of

$$\forall u \in \mathcal{U}_x : \; \pi_x(u) \;:=\; \Pi_x(\{u\}), \tag{28}$$

$$\forall A \in 2^{\mathcal{U}_x} : \; \Pi_x(A) \;:=\; \max_{u \in A} \{\pi_x(u)\}, \tag{29}$$

[13]. Equations (28) and (29) make sense from a semantical point of view, as well. $\Pi_x(A)$ with $A \subset \mathcal{U}_x$ quantifies the possibility that the real value of the variable x is contained in the set A. Therefore dealing with normalized possibility distributions means dealing with possibility measures at the same time. π_x and the corresponding Π_x carry the same information. And any *consonant* information, i.e. any information that corresponds to a *consonant* body of evidence, is completely described by an unequivocally determined possibility measure.

In order to find a dual relationship between σ–distributions and a corresponding fuzzy measure, we recapitulate the concept of a *body of evidence*:

Definition 16 *A function* $m_x : 2^{\mathcal{U}_x} \to [0,1]$ *is called a* body of evidence (basic assignment) *if*

$$m_x(\emptyset) = 0 \textit{ and } \sum_{A \subseteq \mathcal{U}_x} m_x(A) = 1.$$

$m_x(A)$ quantifies our degree of belief that the real value of variable x is contained in the set $A \subseteq \mathcal{U}_x$, but not in any subset $A' \subsetneq A$.

So, assuming an appropriate scaling factor $s \in (0,1]$, from a semantical point of view we may establish the correspondence

$$\forall u \in \mathcal{U}_x : m_x(\{u\}) := s \cdot \sigma_x(u), \tag{30}$$

where the factor s guarantees, that definition 16 does not get violated. We propose to choose s such that

$$\sum_{u \in \mathcal{U}_x} s \cdot \sigma_x(u) = \max_{u \in \mathcal{U}_x} \{\sigma_x(u)\} \tag{31}$$

$$\Leftrightarrow \quad s := \frac{\max_{u \in \mathcal{U}_x} \{\sigma_x(u)\}}{\sum_{u \in \mathcal{U}_x} \sigma_x(u)}, \tag{32}$$

because we feel that the maximal support value taken by σ_x should serve as a measure to evaluate the degree of certainty inherent in σ_x. This is, whenever $\tau^* := \max_{u \in \mathcal{U}_x} \{\sigma_x(u)\}$ is less than 1, there is no value $u \in \mathcal{U}_x$ that is fully supported. So in terms of the basic assignment there is still some evidence $1 - \tau^*$ to be distributed amongst the subsets of $\mathcal{U}_x$. And regarding the principle of minimal completeness once again, we know where to put this remaining evidence grade: in order to assume no information that is not contained within σ_x we have to assign $1 - \tau^*$ to the trivial assumption $x \in \mathcal{U}_x$. We are totally ignorant about this remaining evidence.

So all in all we propose a semantically reasonable way to transform a given σ–distribution σ_x into a corresponding body of evidence m_x by setting

$$\tau^* := \max_{u \in \mathcal{U}_x} \{\sigma_x(u)\}, \; s := \frac{\tau^*}{\sum_{u \in \mathcal{U}_x} \sigma_x(u)} \tag{33}$$

and

$$m_x(A) := \begin{cases} s \cdot \sigma_x(u) & \text{if } A = \{u\}, u \in \mathcal{U}_x, \\ 1 - \tau^* & \text{if } A = \mathcal{U}_x, \\ 0 & \text{else.} \end{cases} \tag{34}$$

m_x obeys the principle of minimal completeness and in fact represents a body of evidence, since $m_x(\emptyset) = 0$ and

$$\sum_{A \subseteq \mathcal{U}_x} m_x(A) = \sum_{u \in \mathcal{U}_x} m_x(\{u\}) + m_x(\mathcal{U}_x) \tag{35}$$

$$= \sum_{u \in \mathcal{U}_x} s \cdot \sigma_x(u) + 1 - \tau^* \tag{36}$$

$$= \tau^* + 1 - \tau^* = 1. \tag{37}$$

The derived body of evidence m_x can now be transformed into two corresponding fuzzy measures Pl_x and Bel_x that represent exactly the same information on variable x:

$$\forall A \subseteq \mathcal{U}_x : Pl_x(A) := \sum_{B : B \cap A \neq \emptyset} m_x(B) \tag{38}$$

$$Bel_x(A) := \sum_{B \subseteq A} m_x(B) \tag{39}$$

But since the focal elements[6] of m_x are not nested, m_x is not a *consonant* body of evidence like in the possibilistic case. And therefore the derived Pl_x is a *plausibility measure* and the derived Bel_x is a *belief measure*. This is, neglecting the scaling factor s any σ–distribution σ_x may be transformed into a corresponding plausibility or belief measure, respectively. And from our point of view, this scaling factor is completely neglectible, in fact, because we never make use of the absolut degree of support $\sigma_x(u)$. The absolute values of any σ–distribution derived never have any particular meaning. All the defuzzification techniques, for instance, yield the same result applied to any $s \cdot \sigma_x$, no matter which value in $(0, 1]$ is assigned to s. In fact, the range $[0, 1]$ of possibility and σ–distributions is just an ordinal scale [11]. This is also verified by the observation, that there is only minimization and maximization used to manipulate possibility or support degrees.

We conclude this section by repeating its main results: Any possibility distribution corresponds to a fuzzy measure with a consonant body of evidence, namely a possibility measure and its dual necessity measure. Neglecting an appropriate scaling factor, any σ–distribution corresponds to a fuzzy measure with a dissonant body of evidence, namely a plausibility measure and its dual belief measure. This is, possibility distributions and σ–distributions represent different types of knowledge.

[6] *Focal elements* of m_x are all the subsets $A \subseteq \mathcal{U}_x$ with $m_x(A) > 0$.

7 Soundness and Completeness

Possibility distributions correspond to consonant bodies of evidence. σ–distributions correspond to dissonant bodies of evidence. Within the context of π–reasoning, to get new information means to get new arguments *against* distinct assumptions, i.e. the corresponding possibility distribution gets smaller (cf. *reduction type inference*, [15]). Within the context of σ–reasoning, to get new information means to get new arguments *in favour of* distinct assumptions, i.e. the corresponding σ–distribution gets bigger (cf. *expansion type inference*, [15]). All these observations endorse the expectation, that π– and σ–reasoning show complementary behaviour w.r.t inconsistent and incomplete knowledge.

And in fact, more closely examining both approaches we discover this very important difference. While π-reasoning is very appropriate to dealing with incomplete knowledge, contradictions within the given information lead to useless conclusions telling us that (almost) *nothing* is possible. If our rule base is incomplete, there is no problem: we will just get an answer π'_y that considers too many elements of $\mathcal{U}_y$ to be possible. But if our rule base is inconsistent, it may happen that there is an argument against any element of the output universe and so no value $v \in \mathcal{U}_y$ seems possible at all. We expect a *sound* fuzzy rule base to transform any normalized input into a likewise normalized output, i.e. whenever the input is sound, the output should be sound, as well. We define

Definition 17 *In the context of π–reasoning, a fuzzy rule base $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ is called ϵ–*sound *iff for the corresponding meta rule $\pi^{\Rightarrow}_{x,y}$ (see equation (14)) it holds*

$$\forall u \in \mathcal{U}_x : \max_{v \in \mathcal{U}_y} \{\pi^{\Rightarrow}_{x,y}(u,v)\} \geq \epsilon.$$

It is very easy to show, that using this definition a ϵ–sound fuzzy rule base guarantees to transform any normalized input π'_x into an output π'_y, that considers at least one element $v \in \mathcal{U}_y$ to be possible with degree $\pi'_y(v) \geq \epsilon$.

σ–reasoning, on the contrary, is perfectly suited to handle inconsistent pieces of information, but possesses distinct disadvantages applied to incomplete knowledge. If the rule base is not sound, we will get an answer σ'_y that just supports too many elements of $\mathcal{U}_y$. No problem. But if our rule base is incomplete, there may be inputs that do not infer any output at all. We expect a *complete* fuzzy rule base to transform any normalized input into a likewise normalized output. We define

Definition 18 *In the context of σ–reasoning, a fuzzy rule base $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ is called* ϵ–complete *iff for the corresponding meta rule $\sigma^{\Rightarrow}_{x,y}$ (see equation (26)) it holds*

$$\forall u \in \mathcal{U}_x : \max_{v \in \mathcal{U}_y} \{\sigma^{\Rightarrow}_{x,y}(u,v)\} \geq \epsilon.$$

Again, using definition 18 we are sure that an ϵ–complete fuzzy rule base transforms an arbitrary input σ'_x into an output σ'_y that supports at least one $v \in \mathcal{U}_y$ with degree $\sigma'_y(v) \geq \epsilon$.

This is, given a fuzzy rule base $[\tilde{A}_i \Rightarrow \tilde{B}_i]$ we can first calculate the corresponding possibilistic representative, i.e. the meta rule $\pi^{\Rightarrow}_{x,y}$, to check how sound the given rule base is. In the second step, we calculate the representative in terms of σ–distributions, i.e. the meta rule $\sigma^{\Rightarrow}_{x,y}$, to evaluate the completeness of the knowledge at hand. And comparing both results we can decide which approach, π–reasoning or σ–reasoning, is more suited to be applied to the rule base.

8 Conclusion

In this paper we introduce the concept of σ–distributions and σ–reasoning. Assuming nothing but the simple semantics of σ–distributions we infer the most common fuzzy control mechanism, MAMDANI's standard fuzzy control (SFC) [10]. This is a very important result, since, at least from our point of view, the theoretical background of SFC was not very convincing. SFC worked in many applications, but there was no completely satisfying explanation, why and when SFC should be used. With the derivation above we present a convincing framework to solve this problem: whenever we interpret the fuzzy predicates given as σ–distributions, SFC constitutes the one and only mechanism to be used. The result derived is also important from a designer's point of view: At first, he now has very simple and concise semantics at hand how to interpret the fuzzy sets involved in SFC design. Furthermore, he is also delivered from selecting a distinct inference mechanism from the vast amount of proposed ones. He does not have to worry about the appropriate implication relation or the use of different s– and t–norms to implement s–t composition or to combine several results. But we hope that this main result will not only help to design lots of new successful applications but will also convince all those researchers still questioning the import of SFC because of its former heuristic touch.

Furthermore, we are able to show, that σ–distributions correspond at least qualitatively to a dissonant bodies of evidence and may therefore be transformed to corresponding belief or plausibility measures.

With σ–reasoning being a concept dual to the well known possibilistic approach to approximate reasoning (π–reasoning) [4, 14], we now have two complementary mechanisms at hand dealing with fuzzy knowledge. We define the notions of ϵ–*soundness* and ϵ–completeness of a fuzzy rule base, with ϵ–soundness belonging to π–reasoning and ϵ–completeness belonging to σ–reasoning. An analysis of the given knowledge base w.r.t. these definitions will help the designer to decide which approach to use, since they show complementary behaviour w.r.t. these properties. SFC based on σ–reasoning, for instance, requires complete, but copes with inconsistent knowledge. Therefore a fuzzy rule base used in SFC should be ϵ–complete with ϵ close to 1. By inspecting successful SFC applications this theoretical result is strongly confirmed.

A closer examination of ϵ–soundness and ϵ–completeness needs to be done in the future. Furthermore, π– and σ–reasoning seem to be very closely related to DUBOIS' and PRADE's notion of *certainty* and *possibility qualification* [3, 6, 5, 7][7]. In order to get an even better understanding of π– and σ–reasoning we plan to examine this relation more closely.

Another important topic for forthcoming investigations is how to transform possibility distributions to σ–distributions and vice versa by examining the measure theoretic background. A semantically correct transformation between both approaches would allow to establish huge knowledge bases built by several modules with different characteristics.

Acknowledgements

The author wants to thank Martin Spott for many exciting and fruitful discussions and his excellent work on possibility distributions.

References

[1] D. Driankov, H. Hellendoorn, and M. Reinfrank. *An Introduction to Fuzzy Control.* Springer, Berlin, 1993.

[2] D. Dubois and H. Prade. The principle of minimum specificity as a basis for evidantial reasoning. In B. Bouchon and R. R. Yager, editors, *Uncertainty in Knowledge-Based Systems (LNCS 286)*, pages 75–84. Springer, Berlin, 1988.

[3] D. Dubois and H. Prade. Basic issues on fuzzy rules and their application to fuzzy control. In D. Driankov, A. Ralescu, P. Eklund, and M. Reinfrank, editors, *IJCAI-91 Workshop on Fuzzy Control*, Sydney, Australia, 1991. Springer.

[7] Attention: π–reasoning corresponds to certainty qualification and σ–reasoning corresponds to possibility qualifiction, surprisingly.

[4] D. Dubois and H. Prade. Fuzzy sets in approximate reasoning, part 1: Inference with possibility distributions. *Fuzzy Sets and Systems*, 40:143–201, 1991.

[5] D. Dubois and H. Prade. Fuzzy rules in knowledge–based systems. In R. R. Yager and Lotfi A. Zadeh, editors, *An introduction to fuzzy logic applications in intelligent systems*. Kluwer, Boston, 1992.

[6] D. Dubois and H. Prade. Possibility theory as a basis for preference propagation in automated reasoning. In *IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE'92)*, San Diego, 1992.

[7] D. Dubois and H. Prade. Fuzzy sets in automated reasoning: problems and methods. In F. Esteva and P. Garćia, editors, *Tecnologías y Lógica Fuzzy (FUZZY'94)*, Blanes, Spain, 1994.

[8] G. J. Klir and T. A. Folger. *Fuzzy Sets, Uncertainty and Information*. Prentice–Hall, Englewood Cliffs, 1988.

[9] C. C. Lee. Fuzzy logic in control systems: Fuzzy logic controller (part I&II). *IEEE Transactions on Systems, Man & Cybernetics*, 20(2):404–435, 1990.

[10] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man–Machine Studies*, 7, 1975.

[11] A. Norwich and I. B. Turksen. Fuzzy expert systems for IE/OR/MS. *Fuzzy Sets and Systems*, 51:1–27, 1992.

[12] W. Pedrycz. Processing in relational structures: Fuzzy relational equations. *Fuzzy Sets and Systems*, 40:77–106, 1991.

[13] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, London, 1976.

[14] M. Spott. Unscharfes Schliessen mit komplexen Wissensbasen. Master's thesis, Universität Karlsruhe (TH), Juli 1994.

[15] I. B. Turksen and Y. Tian. Combination of rules or their consequences in fuzzy expert systems. *Fuzzy Sets and Systems*, 58:3–40, 1993.

[16] Z. Wang and G. J. Klir. *Fuzzy Measure Theory*. Plenum Press, New York, 1992.

[17] J. Weisbrod. On fuzzy implication relations. *Fuzzy Sets and Systems*, 67:211–219, 1994.

[18] R. R. Yager. An introduction to applications of possibility theory. *Human Systems Management*, 3:195–227, 1983.

[19] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[20] Lotfi A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on System, Man, and Cybernetics*, 3:28–44, 1973.

[21] Lotfi A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 8,9:199–249, 301–357, 43–80, 1975.

[22] Lotfi A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

```
@INCOLLECTION{weisbrod95b,
      AUTHOR = {Weisbrod, J.},
      TITLE = {Fuzzy Control Revisited --- Why Is It Working?},
      BOOKTITLE = {Advances in Fuzzy Theory and Technology, Vol.\ III},
      editor = {Wang, P. P.},
      PUBLISHER = {Bookwrights},
      ADDRESS = {Durham (NC)},
      YEAR = {1995},
      PAGES = {219-244}
}
```

A short version of this paper is:

```
@INPROCEEDINGS{weisbrod94c,
      AUTHOR = {Weisbrod, J.},
      TITLE = {Fuzzy Control Revisited --- Why Is It Working?},
      BOOKTITLE = {Joint Conference on Information Sciences / Third
                   Annual Conf.\ on Fuzzy Theory and Technology},
      ADDRESS = {Pinehurst (NC), USA},
      PAGES = {357--360},
      YEAR = {1994}
}
```