Techniques and Applications of Control Systems Based on Knowledge-based Interpolation

Frank Klawonn
Department of Electrical Engineering and Computer Science
FH Ostfriesland, University of Applied Science
Constantiaplatz 4
D-26723 Emden, Germany
e-mail klawonn@et-inf.fho-emden.de

Rudolf Kruse
Department of Computer Science
Otto-von-Guericke University
Universitätsplatz 2
D-39106 Magdeburg, Germany
e-mail kruse@iik.cs.uni-magdeburg.de

Abstract

Fuzzy control was established as an alternative control method when it is difficult to develop a suitable mathematical model of the process, but expert knowledge in the form of vague rule is available. Although the principal idea of a fuzzy set as a model of a vague linguistic expression is very appealing, a naive approach to fuzzy sets can cause tedious problems. Without a concrete semantics for fuzzy sets the design of a fuzzy controller can end up in a trial and error experiment with a large number of parameters and options.

In this paper we review an approach to fuzzy control that interprets fuzzy sets as vague values in a vague environment. The vague environment is characterised by a scaling function that describes how sensitive the process reacts when a certain value is slightly changed. We also discuss a regression technique based on the concept of vague environments, enabling to construct a fuzzy controller from data.

1 Introduction

When L.A. Zadeh introduced the notion of a fuzzy set in his seminal paper [31] in 1965, his intention was the development of a concept for the representation of the vagueness inherent in common linguistic statements like *The temperature should be kept low, high pressure should be avoided* etc. Of course, at that time probabilistic, statistic and stochastic models were already well known and applied in many fields. About the same time as Zadeh introduced fuzzy sets, R.E. Moore published his book on interval analysis [25]. Handling imperfect knowledge and information is the common aim of probabilistic models, interval analysis and fuzzy sets.

However, these paradigms aim at modelling different aspects of imperfect knowledge. Probability is usually understand as the uncertainty whether a certain (well-defined) event occurs. An event can for instance be real number, the outcome of a measuring experiment. Interval analysis is concerned with imprecision, i.e., instead of crisp values intervals are considered without making further assumptions on the specific number in a considered interval. The imprecision itself is considered to be precise, because the interval boundary are assumed to be exact.

Vague values and vague intervals build the underlying interpretation of fuzzy sets in many applications. Linguistic concepts like *slow*, *young* or *small* incorporate vagueness in the sense of a valuated imprecision, meaning that these concepts do not represent crisp values, but ranges or intervals with boundaries that cannot be specified exactly. This viewpoint clarifies that when dealing with vagueness we have to explain how we want to model values or intervals with non-sharp boundaries.

In later years of fuzzy set theory L.A. Zadeh introduced the notion of *computing with words* based on the perception that human thinking is not based on handling numbers, but on operating with linguistic concepts. Nevertheless, for computers processing numbers is essential, starting with input data in the form of real numbers. Thus fuzzy set theory tries to build a bridge between human-like thinking and processing on computers.

The elementary concept in fuzzy set theory is the notion of membership degree that allows for graded memberships of elements to sets. This idea can be used to model vague concepts since it enables us to assign a (vague) property like *fast* to an object like a *car* to a certain degree. Membership degrees as a generalisation of the crisp membership degrees 0 and 1 are often considered as elementary concepts in fuzzy set theory. However, without providing an explanation of what the concrete meaning of a membership value is, there are no unique canonical operations for handling fuzzy sets and a naive approach can lead to contradictions.

For fuzzy systems the meaning of the membership degree is a crucial point. In [4] three different semantics for fuzzy sets are mentioned:

- Preference: The unit interval is considered as nothing more than a linear ordering. The membership degrees are used to formulate flexible constraints in the sense that besides the values zero (forbidden) and one (completely allowed) intermediate values are admitted [2].
- Uncertainty in the sense of possibility: Fuzzy sets are considered as derived concepts from possibility or necessity measures [3, 5, 22, 35].
- Similarity, indistinguishability, indiscernability etc.: These properties are understood in a gradual way and a fuzzy set is for instance induced by one element as the (fuzzy) set of those elements that are indistinguishable from the considered element [1, 32].

Maybe a fourth interpretation should also be mentioned:

- Partial contradiction or partial inconsistency: This interpretation was motivated in [26] and further pursued in [13] on the basis of the Ulam game [29] which deals with an unreliable information source.

In this paper we concentrate on the interpretation of fuzzy sets from the viewpoint of similarity, indistinguishability, or indiscernability.

The paper is organised as follows. Section 2 introduces the notion of vague environments and explains how fuzzy sets can be viewed as points or magnitudes in vague environments. Section 3 explains fuzzy control from the viewpoint of vague environments and the consequences for fuzzy controllers. Finally we use the concept of vague environments to describe a regression technique for learning a fuzzy controller from data in Section 4.

2 Fuzzy Sets, Vague Environments, and Indistinguishability

Many authors distinguish between a fuzzy set M and its associated membership function μ_M. However, they only use M as a name for μ_M and operate always with μ_M. Therefore, we do not distinguish between M and μ_M here. A fuzzy set is a mapping $\mu : X \to [0,1]$ from a universe of discourse X to the unit interval. $\mu(x) \in [0,1]$ is considered as the membership degree of the object or element $x \in X$ to the fuzzy set μ. In many application the universe of discourse X is a real interval, the real numbers, sometimes also a Cartesian product of real intervals. When looking at applications, especially in fuzzy control, one can see that not arbitrary fuzzy sets or membership functions $\mu : \mathbb{R} \to [0,1]$ are used. For instance no one would really want to consider a fuzzy set on the unit interval like

$$\mu(x) = \begin{cases} x & \text{if } x \text{ is a rational number} \\ 0 & \text{otherwise.} \end{cases}$$

Fuzzy sets on real intervals as real-valued functions are almost always 'well-behaved' in the sense that they are continuous and unimodal (having only one local maximum or a range with maximal membership degree). This is of course reasonable, but the pure mathematical definition of a fuzzy set does require any such property.

As we already mentioned earlier, for a consistent handling of fuzzy sets a concrete interpretation of the membership degrees is essential. So in the following we explain how fuzzy sets can be seen as induced concepts, when we start with the elementary and canonical notion of distance.

In engineering applications we have in general to deal with real-valued measurements and control actions in quantified form. We should be aware of the fact that the involved real numbers can never be exact. Of course, in many applications the inexactness is small enough so that we do not have to worry about it.

In the following we will provide a model that is able to represent this inexactness in order to handle problems connected to this phenomenon. Two different forms of inexactness can be distinguished:

- *enforced inexactness* of measurement and control values which is caused by the limited precision of measuring or other instruments or by properties of the physical environment which make an exact measurement impossible.
- *intended imprecision* where we are not interested in arbitrary exactness or where it even does not make sense. As an example consider the room temperature. A difference of 0.0000001°C of the temperature is neither for a human being of interest nor should it influence the air conditioning system. Since human beings do you usually not think in terms of exact numbers but more in the sense of magnitudes, intended imprecision is very much in the spirit of computing with words.

A very simple model of the above mentioned phenomena of inexactness identifies values whose distance is less than an error- or tolerance bound $\varepsilon > 0$. This identification can cause problems since it does not satisfy the law of transitivity, i.e., although x_1 and x_2 as well as x_2 and x_3 are identified according to $|x_1 - x_2| \leq \varepsilon$ and $|x_2 - x_3| \leq \varepsilon$, it is possible that x_1 and x_3 should not be identified due to $|x_1 - x_3| > \varepsilon$. The following situation illustrates this non-transitivity property. The decision to buy a certain luxurious car does in general not depend on an increase of the price of 1$. But it is of course not allowed to iterate this argument, otherwise we would accept any price.

Because of this non-transitivity we cannot define adequate equivalence classes of indistinguishable or identifiable numbers. Enforcing artificial equivalence classes by partitioning the real numbers into disjoint intervals of a certain length and identifying values that fall in the same interval leads automatically to an incoherent treatment of values that are near the boundary of an interval.

Another question is whether we can specify an appropriate tolerance bound ε so that we consider exactly those values as indistinguishable whose difference is less than ε. Choosing ε

too small leads to very tedious and inefficient model that is difficult to handle. If ε is too large the model becomes too rough. Therefore we do not restrict ourselves to just one value of ε but to an interval of possible values for ε. We assume that the interval of possible values for ε is the unit interval. Although this might look like an arbitrary choice, we will see later on when we introduce the concept of scaling that it is sufficient to concentrate on the unit interval.

A very simple approach derived from the above consideration is the following. We replace each exact real numbers (that do actually not come from exact measurements or specifications) by the (fuzzy) set of indistinguishable values. We interpret indistinguishability as the dual concept to distance. Therefore, we define the degree of indistinguishability or similarity between two real number as 1 minus their absolute value of their difference. In order to avoid negative degrees of indistinguishability, we define a fuzzy equivalence relation $E(x,y) = 1 - \min\{|x-y|, 1\}$. In this way, a real is indistinguishable from itself to the degree 1 (completely indistinguishable) and indistinguishable to the degree 0 (completely distinguishable) from all real numbers that differ more than 1 from the considered number.

In the way a real value $x_0 \in \mathbb{R}$ induces the fuzzy set of all real numbers that are indistinguishable from x_0 by

$$\mu_{x_0}(x) \;=\; E(x_0, x) \;=\; 1 - \min\{|x_0 - x|, 1\}.$$

It is quite remarkable that this fuzzy set μ_{x_0} representing the magnitude or approximate value x_0 has a symmetric triangular membership function with its maximum at x_0 and reaching the membership degree at $x_0 - 1$ and $x_0 + 1$. So we have an interpretation and justification for the use of such triangular membership functions, not just because of their simplicity.

However, this very simple approach is too restrictive to be really useful for applications, since it neglects certain important aspects by just defining indistinguishability on the basis of the canonical metric on the real numbers.

The answer to the question whether two values should be considered more or less as the same magnitude is of course problem dependent. But it also depends on the measurement unit. For keeping an aeroplane on a certain flight route a distance of less than three feet is not of importance so that we can identify positions (coordinates) whose distance is less than $\varepsilon = 3$ (feet). But we must not stick to this tolerance bound, when we measure in miles instead of feet.

The same situations appears when we consider temperatures in Celsius or in Fahrenheit. For this reason we allow to introduce a scaling factor $c \geq 0$ [10] and consider two values as ε-distinguishable if their distance times c is greater than ε. So when thinking in terms of Celsius instead of Fahrenheit we have to take a scaling factor $c = 1.8$ into account according to the transformation formula $F = 1.8C + 32$ between Fahrenheit and Celsius.

Let us assume that we have to deal with measurements in the interval $X = [a,b]$. Introducing a scaling factor corresponds to a linear transformation of the interval $[a,b]$ to another interval of length $c(b-a)$, say to the interval $[0, c \cdot (b-a)]$. The transformation is determined by

$$t_c : [a,b] \to [0,\infty), \qquad x \mapsto c \cdot (x-a). \tag{1}$$

The distinguishability of two values $x_1, x_2 \in [a,b]$ is now determined on the basis of the distance of the values, i.e., $|t_c(x_1) - t_c(x_2)|$.

The use of a single scaling factor c enables us to overcome the problem of different scalings as in the case of Fahrenheit and Celsius. However, we are not able to model the fact that a measurement instrument might provide quite precise values in a certain range whereas out of this range the measured values are less reliable. This phenomenon corresponds to enforced inexactness mentioned above. Also in the case of intended imprecision we might wish to distinguish between values in a certain range very carefully, but for other ranges we are not interested in precise values. In order to solve this problem of differing precision for different ranges we introduce varying scaling factors for the ranges. A scaling factor $c > 1$ implies a weak indistinguishability (strong distinguishability) for values in the corresponding range, whereas

error e	scaling factor	interpretation
$-10 > e$	0	don't care about the concrete value of the error, just counteract strongly
$-10 \geq e > -5$	0.5	still a large error, but a little bit care has to be taken
$-5 \geq e \leq 5$	1.5	very sensitive control is necessary to avoid strong overshoots
$5 < e \leq 10$	0.5	same as in the range $-10 \geq e > -5$
$10 < e$	0	same as in the range $e < -10$

Table 1: Scaling factors for the room temperature.

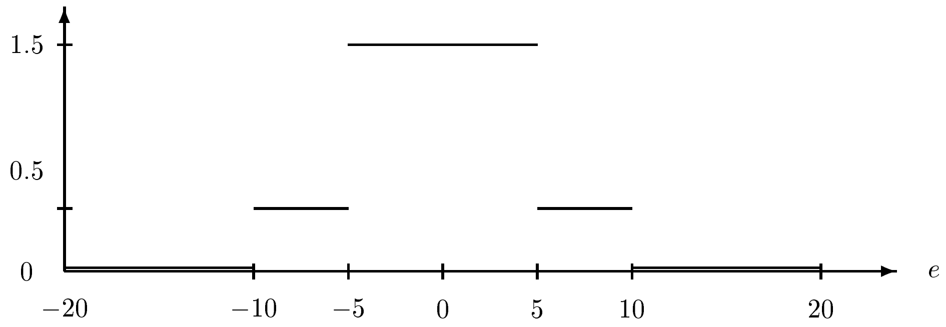

Figure 1: The scaling factor function for the error value.

a scaling factor $c < 1$ leads to a strong indistinguishability. A typical examples for varying indistinguishability for different ranges can be experienced in many control applications. When the system to be controlled is very far away from the desired set value, then strong control actions are necessary. In this case the concrete value of the error – the difference between the actual value and the desired set value – is not very important. It is sufficient just to know in which direction the system is out of order. Therefore, we do not intend to distinguish between measurements values in this range. The situation is completely different when the system has almost reached the desired set value. Then careful and very sensitive control actions have to be taken. It is very important to know whether the error is positive or negative. This means that we distinguish very carefully in the range where the error is near zero.

A possible choice of the scaling factors for this problem is shown in Table 1.

The great scaling factor 1.5 for the range of around zero, i.e. for the error values between -5 and -5, indicates that these values should be distinguished very carefully in order to give a sensitive control action. For the ranges of a medium sized absolute value of the error between 4 and 10 it is important to distinguish between these error values, but since we are still far away from the set point we do not have to be very sensitive. For absolute error values greater than 10 we have to react with the highest possible (positive or negative) value for the control action no matter whether the absolute value of the error is 12 or 19.

Let us assume that $X = [-20, 20]$ is the set of possible error values. The function $c : X \to [0, \infty)$, assigning to each error value the corresponding scaling factor, is shown Figure 1. The corresponding transformation induced by these scaling factors is illustrated in Figure 2.

It is easy to check that the piecewise linear transformation in Figure 2 from $X = [a, b] =$

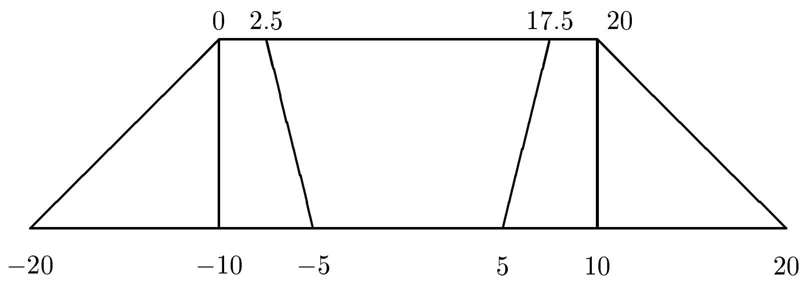

Figure 2: The transformation induced by the scaling function in Figure 1.

$[-20, 20]$ to $[0, 20]$ can be computed by

$$t_c : [a, b] \to [0, \infty), \qquad x \mapsto \int_a^x c(s)ds \tag{2}$$

where the function c is given by

$$c : X \to [0, \infty), \qquad s \mapsto \begin{cases} 0 & \text{if} \quad -20 \leq s < -10 \\ 0.5 & \text{if} \quad -10 \leq s < -5 \\ 1.5 & \text{if} \quad -5 \leq s < 5 \\ 0.5 & \text{if} \quad 5 \leq s < 10 \\ 0 & \text{if} \quad 10 \leq s < 20. \end{cases} \tag{3}$$

If we choose a tolerance bound $\varepsilon = 0.5$ then the error values 0.8 and 1.2 are distinguishable (w.r.t. this tolerance bound) whereas 9 and 11 are indistinguishable. By Equation (2) the transformed values for 0.8, 1.2, 9, and 11 are 11.2, 11.8, 19.5, and 20, respectively.

Note that Equation (2) coincides with Equation (1) when we choose a constant scaling function c.

The scaling function in Figure 1 reflects the idea that we distinguish values near the optimal set value very carefully, whereas the distinguishability decreases the farther away we go from this value. The piecewise linear function was only chosen to elucidate the principle of different scaling factors and to have a simple transformation function. In the most general case we associate with each value x of our set $X = [a, b]$ a scaling factor $c(x) \geq 0$. The function c has not to be piecewise linear. All we have to assume is that c is integrable. For the transformation induced by such a general scaling function Equation (2) is still valid. To understand this, we assume that there is an integrable scaling function $c : [a, b] \to [0, \infty)$. In the neighbourhood of the point x_0 the transformed (directed) distance $\delta_c^{\mathrm{dir}}(x_0, x)$ between x_0 and a point x very near to x_0 should be approximately $c(x_0) \cdot (x - x_0)$, i.e. $\delta_c^{\mathrm{dir}}(x_0, x) \approx c(x_0) \cdot (x - x_0)$, or

$$\frac{\delta_c^{\mathrm{dir}}(x_0, x)}{x - x_0} \approx c(x_0)$$

For $x \to x_0$ we assume

$$\lim_{x \to x_0} \frac{\delta_c^{\mathrm{dir}}(x_0, x)}{x - x_0} = c(x_0)$$

Thus we have

$$\frac{\partial \delta_c^{\mathrm{dir}}(x_0, x)}{\partial x} = c(x_0)$$

so that we obtain

$$\delta_c^{\text{dir}}(x_0, x) = \int_{x_0}^{x} c(s)ds + \text{const.}$$

Since $\delta_c^{\text{dir}}(x_0, x_0) = 0$, we conclude const $= 0$ and obtain

$$\delta_c(x_1, x_2) = \left|\delta_c^{\text{dir}}(x_1, x_2)\right| = \left|\int_{x_1}^{x_2} c(s)ds\right|. \tag{4}$$

x_1 and x_2 are considered to be ε-distinguishable with respect to the scaling function c if their 'transformed distance' $\delta_c(x_1, x_2)$ is greater than ε.

We now turn to the problem of representing a *vague environment* that is characterised by a distance function δ_c of the above mentioned type. We do not consider only one fixed value ε, but a whole set of values for ε, each of them leading to a different ε-distinguishability. We consider all numbers from the unit interval as possible values for ε. If one would prefer to have a smaller or larger interval as possible values for ε, this can be amended by an appropriate choice of the scaling function c. If for example the scaling function c is replaced by the scaling function $\hat{c} = \lambda \cdot c$ then ε-distinguishability with respect to c corresponds to (ε/λ)-distinguishability with respect to $\hat{c}$. In this sense allowing all values from the unit interval for ε covers already the most general case.

For each $\varepsilon \in [0, 1]$ we associate with the value $x_0 \in X$ all values $x \in X$ which are not ε-distinguishable from x_0 (with respect to the scaling function c), i.e. the set

$$S_{x_0,\varepsilon} = \{x \in X \mid \delta_c(x, x_0) \leq \varepsilon\}. \tag{5}$$

A more convenient representation of this family of sets is described by the mapping

$$\mu_{x_0} : X \to [0, 1], \quad x \mapsto 1 - \min\{\delta_c(x, x_0), 1\}, \tag{6}$$

so that we have

$$S_{x_0,\varepsilon} = \{x \in X \mid \mu_{x_0}(x) \geq 1 - \varepsilon\}.$$

$\mu_{x_0}(x)$ can be interpreted intuitively as the degree to which x can be identified with x_0. Therefore we can understand μ_{x_0} as the fuzzy set of values that are indistinguishable to x_0. Note that in the most simple case where we have the same scaling factor $c > 0$ for all $x \in X$, i.e., a constant scaling function, we obtain a triangular membership function with slope c taking its maximum at x_0 as the fuzzy set μ_{x_0} which represents the value x_0 in the vague environment X.

Note that the α-cut $\{x \in X \mid \mu_{x_0} \geq \alpha\}$ of the fuzzy set μ_{x_0} is equal to the set $S_{x_0,1-\alpha}$ of elements that are $(1 - \alpha)$-indistinguishable from x_0.

Let us return to the vague environment characterised by the scaling function introduced in Equation (3). Figure 3 illustrates the fuzzy sets μ_{x_0} that are associated with the values $x_0 = -10, -8, -5, -3$ in this vague environment, i.e., the fuzzy sets of real numbers that are indistinguishable from these values.

The fuzzy sets in Figure 3 are all of triangular or trapezoidal type. This is not necessarily the case as Figure 4 illustrates where the fuzzy sets μ_{x_0} associated with the values $x_0 = -9$ and $x_0 = -4.5$ are shown. (Note the different scaling on the x-axis in comparison to Figure 3.)

Since the scaling function in Equation (3) is piecewise constant, the fuzzy set μ_{x_0} associated with a value x_0 will always be piecewise linear.

To obtain other shapes for the fuzzy set associated with the value x_0, an appropriate scaling function has to be defined. As an example let us consider a bell shaped fuzzy set of the form

$$\mu : \mathbb{R} \to [0, 1], \quad x \mapsto \exp\left(-\frac{1}{2}\left(\frac{x - x_0}{\sigma}\right)^2\right). \tag{7}$$

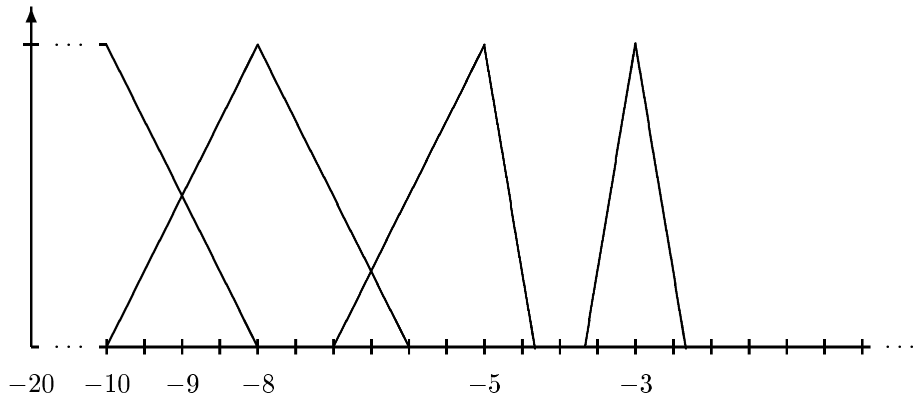

Figure 3: The fuzzy sets μ_{x_0} associated with the values $x_0 = -10, -8, -5, -3$ in the vague environment induced by the scaling function in Equation (3).

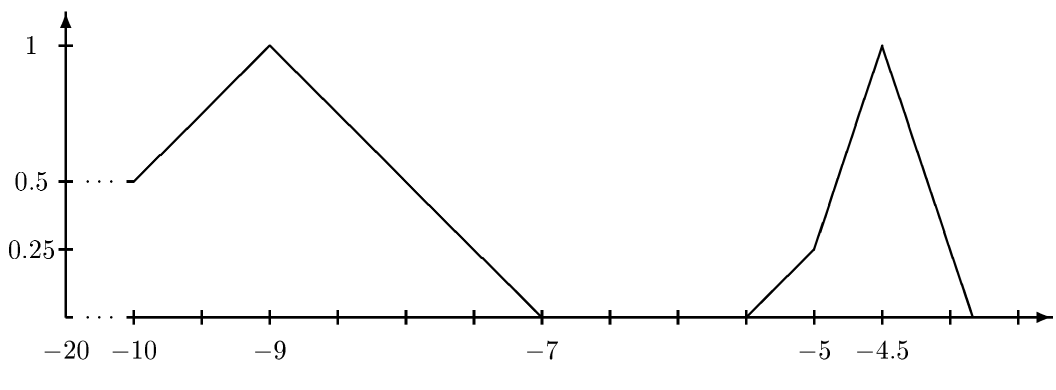

Figure 4: The fuzzy sets μ_{x_0} associated with the values $x_0 = -9$ and $x_0 = -4.5$ in the vague environment induced by the scaling function in Equation (3).

Choosing

$$c : \mathbb{R} \to [0,\infty), \qquad x \mapsto \frac{|x-x_0|}{\sigma^2} \cdot \exp\left(-\frac{1}{2}\left(\frac{x-x_0}{\sigma}\right)^2\right) \tag{8}$$

as the scaling function, we obtain $\mu = \mu_{x_0}$, i.e. the fuzzy set μ represents the value x_0 in the vague environment induced by the scaling function c. This is a direct consequence of Equation (4) for the transformed distance, since (8) is simply the absolute value of the first derivative of (7).

More generally we have the following theorem.

Theorem 1 *Let $\mu : \mathbb{R} \to [0,1]$ be a fuzzy set such that there exists $x_0 \in \mathbb{R}$ with*

(i) $\mu(x_0) = 1$,

(ii) μ is a non-decreasing function on $(-\infty, x_0]$,

(iii) μ is a non-increasing function on $[x_0, \infty)$,

(iv) μ is continuous,

(v) μ is almost everywhere differentiable.

Then there exists a scaling function $c : \mathbb{R} \to [0,\infty)$ such that μ coincides with the fuzzy set μ_{x_0} which is associated with the value x_0 in the vague environment induced by c.

Proof. Choose $c(x) = \left|\frac{d\mu(x)}{dx}\right|$ as the scaling function. $\diamond$

It is obvious, that the reverse of Theorem 1 also holds, which means that, given a scaling function $c : \mathbb{R} \to [0,\infty)$ and a value x_0, then the fuzzy μ_{x_0} associated with x_0 in the vague environment induced by c satisfies conditions (i) – (v) of Theorem 1.

Conditions (i) – (iii) guarantee that the fuzzy set is fuzzy convex (i.e. all its α-cuts are convex), so that it can be considered as the representation of a single value in a vague environment. Fuzzy sets that are not fuzzy convex cannot appear in vague environments when fuzzy sets stand only for single values. It is, of course, possible not just to associate with a single value a fuzzy set in a vague environment, but to associate with any set of values a corresponding fuzzy set by generalising Equations (5) and (6) for a set $M \subseteq X$ by

$$S_{M,\varepsilon} = \{x \in X \mid \exists x_0 \in M : \ \delta_c(x, x_0) \leq \varepsilon\}.$$

and

$$\mu_M : X \to [0,1], \qquad x \mapsto 1 - \min\left\{\inf_{x_0 \in M}\{\delta_c(x,x_0)\}, 1\right\},$$

respectively. Figure 5 illustrates an example for such a fuzzy set associated with the set $M = \{2,4\}$ in the vague environment induced by the constant scaling function $c = 0.5$. This fuzzy set is not fuzzy convex.

Up to now we have only considered fuzzy sets as representations of crisp values in vague environments that were described by scaling functions. In this way a fuzzy partition (set of fuzzy sets) as in Figure 3 is induced by a set of crisp values together with a scaling function. We now turn to the question whether we can provide a scaling function for a given set of fuzzy sets such that the corresponding fuzzy sets can be interpreted as representations of crisp values in the vague environment characterised by the scaling function.

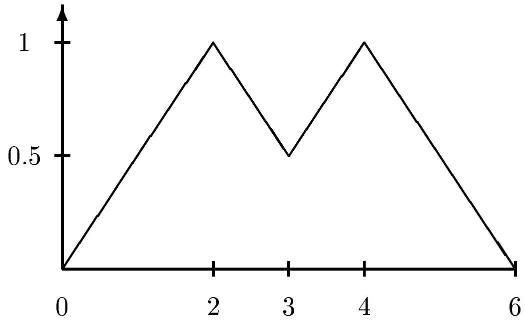

Figure 5: The fuzzy set $\mu_{\{2,4\}}$ in the vague environment induced by the constant scaling function $c = 0.5$.

Theorem 2 *Let $(\mu_i)_{i\in I}$ be an at most countable family of fuzzy sets on $\mathbb{R}$ and let $(x_0^{(i)})_{i\in I}$ be a family of real numbers such that $\mu_i(x_0^{(i)}) = 1$ holds and the conditions (i) – (v) of Theorem 1 are satisfied for all $i \in I$. There exists a scaling function $c : \mathbb{R} \to [0,\infty)$ such that μ_i coincides with the fuzzy set $\mu_{x_0^{(i)}}$ (for each $i \in I$) induced by the value $x_0^{(i)}$ in the vague environment induced by the scaling function c, if and only if*

$$\min\{\mu_i(x), \mu_j(x)\} > 0 \quad \Rightarrow \quad \left|\frac{d\mu_i(x)}{dx}\right| = \left|\frac{d\mu_j(x)}{dx}\right| \tag{9}$$

holds almost everywhere for all $i, j \in I$.

Proof. Assume that (9) is satisfied. Define the scaling function

$$c : \mathbb{R} \to [0,\infty), \qquad x \mapsto \begin{cases} \left|\frac{d\mu_i(x)}{dx}\right| & \text{if } \mu_i(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(2) guarantees that c is well defined almost everywhere. Theorem 1 yields that $\mu_i = \mu_{x_0^{(i)}}$ holds for all $i \in I$. Note that it is sufficient for the proof of Theorem 1 that the scaling function coincides with the derivate of the fuzzy set only on the support of the fuzzy set.

In order to prove the reverse implication, we assume now that there is a scaling function $c : \mathbb{R} \to [0,\infty)$ such that $\mu_i = \mu_{x_0^{(i)}}$ holds for all $i \in I$. Let $i, j \in I$ and let $x \in \mathbb{R}$ with $\min\{\mu_i(x), \mu_j(x)\} > 0$. By definition we have

$$\mu_k(x) \;=\; 1 - \left|\int_{x_0^{(k)}}^{x} c(s)ds\right|$$

for $k \in \{i, j\}$, which implies

$$\left|\frac{d\mu_k(x)}{dx}\right| \;=\; c(x)$$

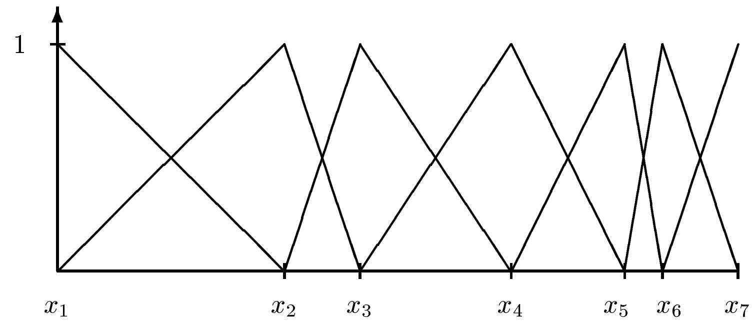

Figure 6: A typical fuzzy partition for which a scaling function can be defined easily.

if μ_k is differentiable at x. Since μ_i and μ_j are almost everywhere differentiable, we obtain

$$\left|\frac{d\mu_i(x)}{dx}\right| \;=\; c(x) \;=\; \left|\frac{d\mu_j(x)}{dx}\right|$$

almost everywhere. ◇

Theorem 2 simply states that we can find a corresponding scaling function for a given fuzzy 'partition' if for each real number $x \in \mathbb{R}$ the absolute value of the slope at x is the same for all fuzzy sets in the fuzzy partition whenever x belongs to the support of the fuzzy set.

A very common type of fuzzy partition is obtained by choosing crisp values $x_1 < x_2 < \ldots < x_n$ and defining the fuzzy set μ_i (for $1 < i < n$) by a triangular membership function which takes its maximum at x_i and reaches the value zero at x_{i-1} and x_{i+1}, respectively. Such a fuzzy partition is illustrated in Figure 6.

For such fuzzy partitions the corresponding scaling function can be defined as the piecewise constant function

$$c(x) \;=\; \frac{1}{x_{i+1} - x_i} \qquad \text{if } x_i < x < x_{i+1},$$

so that the fuzzy sets μ_i represent the fuzzy sets induced by the values x_i in the vague environment induced by the scaling function c.

When we interpret fuzzy sets on the basis of underlying scaling functions, narrow fuzzy sets indicate that in the range of their support we have to be quite careful about the exact value, whereas wider fuzzy sets should be used in regions where the exact value is so important.

The use of scaling functions does not admit arbitrary fuzzy partitions. The fuzzy sets have to be coherent in the sense of Theorem 2, i.e., the absolute values of the first derivative of any two fuzzy sets have to be equal on the intersection of the supports of these fuzzy sets (almost everywhere). This corresponds to the above mentioned observation that the wideness of a fuzzy set characterises the indistinguishability in a region. Therefore, fuzzy sets with different values for the first derivative in the same position would mean that we have low and high indistinguishability in this place at the same time which would make no sense.

So far we have restricted our considerations to a single real valued domain. In general we have to deal with multi-dimensional spaces in control application, the dimension depending on the number of input and output variables. Indistinguishability is of course inherent in multi-dimensional spaces as well. A possible starting point for the $\mathbb{R}^p$ could be the metric induced by a norm, i.e. $\delta(x, y) = \| x - y \|$ for $x, y \in \mathbb{R}^p$. $\| \,.\, \|$ could for instance be the Euclidean norm. However, the concept of scaling cannot be transferred to multi-dimensional spaces in a simple way, because there are infinitely many possible directions in any point and inifinetely many possible candidates as shortest (scaled) ways between two points. We would immediately run into problems of Riemann geometry. Another problem is that humans usually do not think in terms of multi-dimensional spaces. Operating on single variables and combining them later on is a more convenient approach to handle multi-dimensional spaces. Thus we carry out the specification of the vague environment in the single spaces by scaling functions and aggregate them in the product space. There are various possibilities to carry out this aggregation. Here we mention only two. For a more general treatment we refer to [16].

When we want to define the degree of indistinguishability between the two vectors $(x_1, \dots, x_p) \in \mathbb{R}^p$ and $(y_1, \dots, y_p) \in \mathbb{R}^p$, a simple approach is to take the minimum of the indistinguishability degrees in each dimension. So if $\delta_{c_i}(x_i, y_i)$ is the scaled distance in the ith dimension, i.e. the indistinguishability degree between x_i and y_i is $E_i(x_i, y_i) = 1 - \min\{\delta_{c_i}(x_i, y_i), 1\}$, then the indistinguishability degree between the vectors is

$$E\big((x_1, \dots, x_p), (y_1, \dots, y_p)\big) \;=\; \min\{E_i(x_i, y_i)\}. \tag{10}$$

This indistinguishability can be considered as induced by the metric

$$\delta\big((x_1, \dots, x_p), (y_1, \dots, y_p)\big) \;=\; \max\{\delta_{c_i}(x_i, y_i)\},$$

i.e. $E = 1 - \min\{\delta, 1\}$. Using this definition, the indistinguishability of two vectors depends only on the lowest indistinguishability of their components.

An alternative to the minimum is based on the Łukasiewicz-t-norm $\alpha * \beta = \max\{\alpha + \beta - 1, 0\}$, defining the indistinguishability by

$$E\big((x_1, \dots, x_p), (y_1, \dots, y_p)\big) \;=\; \sum_{i=1}^{p} E_i(x_i, y_i) \;-\; (p-1).$$

This indistinguishability can be considered as induced by the metric

$$\delta\big((x_1, \dots, x_p), (y_1, \dots, y_p)\big) \;=\; \sum_{i=1}^{p} \delta_{c_i}(x_i, y_i).$$

This approach takes all the indistinguishabilities in the single dimensions for the aggregation into account.

Nevertheless, in any of these approaches the underlying assumption is that the indistinguishabilities in the single domains can be specified independently. In the following section we will see that this assumption applies to fuzzy control as well, although it usually satisfied in control applications. Consider for instance a controller that uses the error e and the change of the error Δe as inputs. We consider the indistinguishability for the change of the error. If the error is very large, then a strong counteracting control action has to be carried more or less independent of the value of Δe. Therefore we could conclude that we can be content with a high indistinguishability for Δe, since (in the case of a large error) the exact value of Δe is not so important. However, when the error is almost zero, then it really necessary to be very careful about the value of Δe. For the control action it is crucial to distinguish very well at least in the neighbourhood of zero, because we have to know the sign of Δe indicating in which direction the system is changing. These considerations show that the indistinguishability in the domain of Δe depends on the value of the variable e, a phenomenon that we cannot model when we characterise the indistinguishabilities on the single domains by scaling functions. Although the might be an argument against our model and even against fuzzy control, this is not really true as the successful application show. In engineering application we usually have to find a compromise between an almost perfect, but very complicated model with sophisticated parameters and a simplifying approach that is very easy to handle.

3 Application to Fuzzy Control

We now provide a framework for fuzzy control based on the concept of vague environments and indistinguishability and establish the connection to Mamdani's fuzzy control model.

Let us first describe the (simplified) formalisation of the control problem that we want to consider here. We are given n input variables $\xi_1, \ldots, \xi_n$ taking values in the sets $X_1, \ldots, X_n$, respectively. For reasons of simplicity we assume that we have one output- or control variable η with values in the set Y. The problem we have to solve is to find an adequate control function $\varphi : X_1 \times \ldots \times X_n \to Y$, that assigns to each input tuple $(x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n$ an appropriate output value $y = \varphi(x_1, \ldots, x_n) \in Y$.

Before we introduce knowledge-based control making use of the notion of vague environments, we shortly recall how Mamdani's fuzzy control model [24] is defined, since we will see later on that we obtain also Mamdani's model as a result of our approach. The control function φ is specified by k linguistic control rules R_r in the form

$$R_r : \text{ if } \xi_1 \text{ is } A^{(1)}_{i_1,r} \text{ and } \ldots \text{ and } \xi_n \text{ is } A^{(n)}_{i_n,r} \text{ then } \eta \text{ is } B_{i_r} \qquad (r = 1, \ldots, k),$$

where each linguistic term $A^{(1)}_{i_1,r}$, ..., $A^{(n)}_{i_n,r}, B_{i_r}$ is associated with a fuzzy set $\mu^{(1)}_{i_1,r}$, ..., $\mu^{(n)}_{i_n,r}$, μ_{i_r} on $X_1, \ldots, X_n, Y$, respectively. If we are given the input tuple $(x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n$, the output of Mamdani's fuzzy controller is the fuzzy set

$$\mu^{\text{output}}_{x_1,\ldots,x_n} : Y \to [0,1], \qquad y \mapsto \max_{r \in \{1,\ldots,k\}} \left\{ \min\{\mu^{(1)}_{i_1,r}(x_1), \ldots, \mu^{(n)}_{i_n,r}(x_n), \mu_{i_r}(y)\} \right\}$$

on Y. In order to obtain a crisp output value, the fuzzy set $\mu^{\text{output}}_{x_1,\ldots,x_n}$ has to be defuzzified. A very common defuzzification strategy is the centre of area method, but also the mean of maximum- and the max criterion method are applied (see for example [23]). For our purposes, it is sufficient to know that these defuzzification strategies compute a crisp value from a fuzzy set, the exact algorithm for each method is not of importance in this section.

We now come to the presentation of a concept of knowledge-based control based on vague environments, which looks at first glance completely different from Mamdani's model, although there are parallels to the rationale behind Zadeh's compositional rule of inference [33, 34]. But it turns out that the same computations are carried out. Thus we are able to provide a reasonable semantics for Mamdani's model.

The first step in the design of a controller based on vague environments is the specification of appropriate scaling functions $c_1, \ldots, c_n, c$ on the sets $X_1, \ldots, X_n, Y$, respectively. These scaling functions are intended to model the indistinguishability or similarity of values as described in the previous section and induce the fuzzy equivalence relations

$$E_i(x_i, x_i') \;=\; 1 - \min\left\{ \left| \int_{x_i}^{x_i'} c_i(s)ds \right|, 1 \right\}$$

and

$$F(y, y') \;=\; 1 - \min\left\{ \left| \int_{y}^{y'} c(s)ds \right|, 1 \right\}$$

on the spaces X_i and Y, respectively.

Remember that there are two different concepts of indistinguishability. In fuzzy control we mainly have to deal with intended imprecision, which is not enforced by difficulties in measuring exact values, but which is intended to model the fact that arbitrary precision is not needed. Later on we can make use of this fact, since it will be sufficient to specify controller outputs only for certain 'typical' values. Taking the (intended) indistinguishability into account, we can extend this partially defined control function to a fully defined one.

Scaling functions are very appealing, since they have a reasonable interpretation. Small scaling factors imply a low distinguishability, meaning that in this area the control action changes only slowly with varying input values. A greater scaling factor indicates a high distinguishability, i.e., even small variations of the inputs might lead to greater alterations in the control action.

In the next step a control expert has to provide a set of input-output tuples, i.e., tuples $\big((x_1^{(r)},\ldots,x_n^{(r)}),y^{(r)}\big) \in (X_1\times\ldots\times X_n)\times Y$ $(r=1,\ldots,k)$. The tuple $\big((x_1^{(r)},\ldots,x_n^{(r)}),y^{(r)}\big)$ simply means that $y^{(r)}$ is the appropriate output value for input $(x_1^{(r)},\ldots,x_n^{(r)})$. These k input-output tuples correspond to a partial specification of the control function, since they can be understood as the function

$$\varphi_0 : \big\{(x_1^{(r)},\ldots,x_n^{(r)}) \mid r\in\{1,\ldots,k\}\big\} \to Y, \qquad (x_1^{(r)},\ldots,x_n^{(r)}) \mapsto y^{(r)},$$

which is a partial mapping from $X_1\times\ldots\times X_n$ to Y.

Our task is now to determine an appropriate output value $\varphi(x_1,\ldots,x_n)=y\in Y$ for an arbitrary input $(x_1,\ldots,x_n)\in X_1\times\ldots\times X_n$ from the knowledge given by the indistinguishabilities induced by the scaling functions and the partial control function. We extend the indistinguishabilities given by the scaling functions on the domains $X_1,\ldots,X_n,Y$ to an indistinguishability on the product space $X_1\times\ldots\times X_n\times Y$, say in the form of Equation (10), i.e.

$$E\big((x_1,\ldots,x_n,y),(x_1',\ldots,x_n',y')\big) = \min\{E_1(x_1,x_1'),\ldots,E_n(x_n,x_n'),F(y,y')\}.$$

Now we consider the fuzzy set of points in the product space $X_1\times\ldots\times X_n\times Y$ that are indistinguishable from one of the points $((x_1^{(r)},\ldots,x_n^{(r)}),y^{(r)})$:

$$\mu_{\varphi_0}(x_1,\ldots,x_n,y) = \max_{r\in\{1,\ldots,k\}}\big\{E\big((x_1^{(r)},\ldots,x_n^{(r)},y^{(r)}),(x_1,\ldots,x_n,y)\big)\big\}$$

on $X_1\times\ldots\times X_n\times Y$. To obtain an 'output' for a given input tuple $(x_1,\ldots,x_n)$ we compute the projection of this fuzzy set at $(x_1,\ldots,x_n)$, which leads to the fuzzy set

$$\mu_{\varphi_0}^{(x_1,\ldots,x_n)}(y) = \max_{r\in\{1,\ldots,k\}}\big\{\min\{E_1(x_1^{(r)},x_1),\ldots,E_n(x_n^{(r)},x_n),F(y^{(r)},y)\}\big\} \tag{11}$$

on Y. Remembering that the fuzzy set $E(\,.\,,x_0)$ stands for the points that are indistinguishable from the point x_0 with respect to the vague environment induced by the corresponding scaling function, we can rewrite the fuzzy set (11) in the form

$$\mu_{\varphi_0}^{(x_1,\ldots,x_n)}(y) = \max_{r\in\{1,\ldots,k\}}\big\{\min\{\mu_{x_1^{(r)}}(x_1),\ldots,\mu_{x_n^{(r)}}(x_n),\mu_{y^{(r)}}(y)\}\big\}, \tag{12}$$

since $E_1(x_1^{(r)},\,.\,),\ldots,E_n(x_n^{(r)},\,.\,),F(y^{(r)},\,.\,)$ corresponds to the fuzzy set $\mu_{x_1^{(r)}},\ldots,\mu_{x_n^{(r)}},\mu_{y^{(r)}}$, respectively.

Now we are able to see the connection to Mamdani's model. For our approach we started with the specification of scaling functions on the sets $X_1,\ldots,X_n,Y$ and a partial control mapping $\varphi_0 : X_1\times\ldots\times X_n\to Y$ in the form of the set

$$\big\{\big((x_1^{(r)},\ldots,x_n^{(r)}),y^{(r)}\big) \mid r\in\{1,\ldots,k\}\big\}. \tag{13}$$

Taking into account that due to the indistinguishability characterised by the scaling functions we are working in vague environments, this partial mapping can be interpreted as k control rules of the form

$$\begin{array}{lll} R_r: & \text{if } \xi_1 \text{ is } \textit{approximately } x_1^{(r)} \text{ and } \ldots \text{ and } \xi_n \text{ is } \textit{approximately } x_n^{(r)} & \\ & \text{then } \eta \text{ is } \textit{approximately } y^{(r)} & (r=1,\ldots,k) \end{array} \tag{14}$$

where *approximately* $x_1^{(r)}$, ..., *approximately* $x_n^{(r)}$, *approximately* $y^{(r)}$ is represented by the fuzzy set $\mu_{x_1^{(r)}}, \ldots, \mu_{x_n^{(r)}}, \mu_{y^{(r)}}$, respectively. Taking the above control rules together with these fuzzy sets, we can define a fuzzy controller in the sense of Mamdani and obtain for the input $(x_1, \ldots, x_n)$ the fuzzy set $\mu^{\text{output}}_{x_1,\ldots,x_n}$ as output. This fuzzy set coincides with the fuzzy set $\mu_{\varphi_0}^{(x_1,\ldots,x_n)}$, which is the output derived in our knowledge-based control model based on vague environments. In this way we can translate our control approach to Mamdani's model and obtain in both models the same output (before defuzzification).

The obvious question that turns up is, whether a fuzzy controller in the sense of Mamdani can be translated to a controller based on vague environments. The answer is yes if the fuzzy partitions used in Mamdani's model satisfy the conditions mentioned in Theorem 2

Viewing Mamdani's model in the light of fuzzy equivalence relations provides also explanations for the choice of typical fuzzy partitions. The scaling functions should be chosen depending on how sensitive the process reacts when the corresponding value changes. But how should the interpolation points for the partial control function be chosen? Of course, it might be reasonable to specify as many interpolation points as possible. However, we stick here to the philosophy that the expert tries to define as few interpolation points as are necessary for a satisfactory description of the function. This method frees the expert from specifying redundant knowledge and leads to a very information compressed representation of the function to be interpolated. Let us assume that the output $y_0^{(i)}$ for the imprecisely known input $x_0^{(i)}$ is given. The fuzzy equivalence relation E induced by the scaling function c on X enables us to get information about the output corresponding to the value x, as long as $E(x, x_0^{(i)}) > 0$ holds. Thus the next imprecisely known interpolation points $x_0^{(i-1)}$ and $x_0^{(i+1)}$ should be chosen such that $E(x_0^{(i-1)}, x_0^{(i)}) = 0 = E(x_0^{(i+1)}, x_0^{(i)})$ and $E(x, x_0^{(i)}) > 0$ for all $x_0^{(i-1)} < x < x_0^{(i+1)}$. If we follow this minimality philosophy, we obtain a fuzzy partition from the imprecisely known values $x_0^{(i)}$ that satisfies the condition $\mu_i(x) + \mu_{i+1}(x) = 1$ for all $x_0^{(i)} < x < x_0^{(i+1)}$. Thus we can provide an interpretation for such typical fuzzy partitions in terms of a 'lazy' expert who specifies as few interpolation points as necessary.

In order to elucidate our new theoretical and semantic approach to fuzzy control, we shortly review an application to engine idle speed control [12] in the light of scaling functions. We do not go into the technical details of engine idle speed control, but restrict ourselves to the formal specification of the fuzzy controller and how it can be reformulated in terms of scaling functions.

Two input variables are used for the controller, namely the deviation dREV of the number of revolutions to the target rotation speed of the engine and the gradient gREV of the number of revolutions (to be understood as the difference of numbers of revolutions w.r.t. two measurement points). The change of current dAARCUR for the auxiliary air regulator serves as the output variable to influence the rotation speed.

The fuzzy sets defined on the domains of these variables are shown in Figure 7. The underlying domains are $X^{(\text{dREV})} = [-70, 70]$ (rotations per minute), $X^{(\text{gREV})} = [-40, 40]$ (rotations per minute), and $Y^{(\text{dAARCUR})} = [-25, 25]$, where the latter one is to be interpreted as a linear transformation of the real value of the current change dAARCUR.

The rule base of the fuzzy controller is given in Table 2

For the fuzzy partitions Theorem 2 is applicable so that we obtain the scaling functions

$$c_{\text{dREV}} : X^{(\text{dREV})} \longrightarrow [0, \infty), \qquad x \mapsto \begin{cases} \frac{1}{20}, & \text{if } -70 \le x < -30 \\ \frac{1}{28}, & \text{if } -30 \le x < -2 \\ 0, & \text{if } -2 \le x < 2 \\ \frac{1}{28}, & \text{if } 2 \le x < 30 \\ \frac{1}{20}, & \text{if } 30 \le x < 70 \end{cases}$$

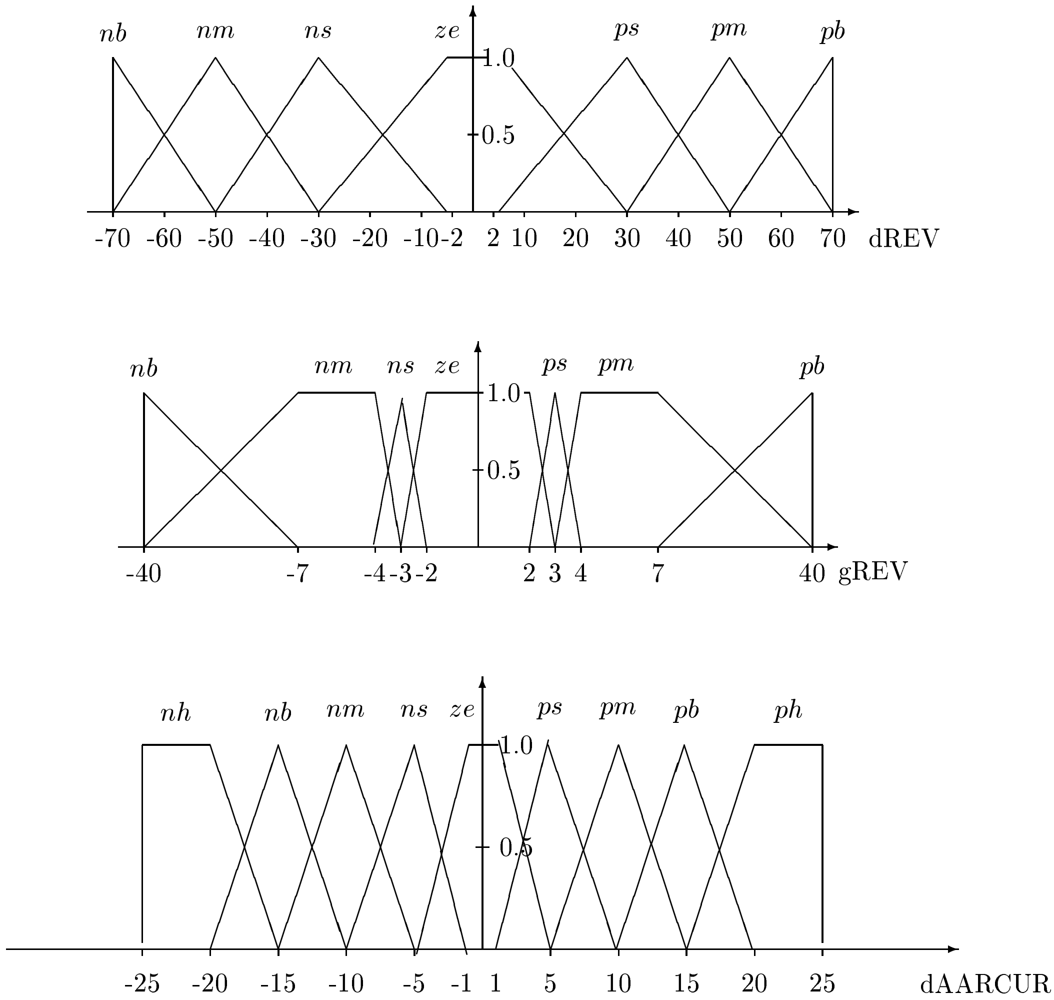

Figure 7: : The fuzzy partitions of for the variables dREV, gREV, and dAARCUR

gREV

dREV	*nb*	*nm*	*ns*	*ze*	*ps*	*pm*	*pb*
nb	*ph*	*pb*	*pb*	*pm*	*pm*	*ps*	*ps*
nm	*ph*	*pb*	*pm*	*pm*	*ps*	*ps*	*ze*
ns	*pb*	*pm*	*ps*	*ps*	*ze*	*ze*	*ze*
ze	*ps*	*ps*	*ze*	*ze*	*ze*	*nm*	*ns*
ps	*ze*	*ze*	*ze*	*ns*	*ns*	*nm*	*nb*
pm	*ze*	*ns*	*ns*	*nm*	*nb*	*nb*	*nh*
pb	*ns*	*ns*	*nm*	*nb*	*nb*	*nb*	*nb*

Table 2: The rules to control the engine idle speed

	nh	*nb*	*nm*	*ns*	*ze*	*ps*	*pm*	*pb*	*ph*
$X^{(\text{dREV})}$	–	−70	−50	−30	0	30	50	70	–
$X^{(\text{gREV})}$	–	−40	−5	−3	0	3	5	40	–
$X^{(\text{dAARCUR})}$	−20	−15	−10	−5	0	5	10	15	20

Table 3: The fuzzy sets corresponding to the points in the vague environments

$$c_{\text{gREV}} : X^{(\text{gREV})} \longrightarrow [0,\infty), \qquad x \mapsto \begin{cases} \frac{1}{33}, & \text{if } -40 \leq x < -7 \\ 0, & \text{if } -7 \leq x < -4 \\ 1, & \text{if } -4 \leq x < -2 \\ 0, & \text{if } -2 \leq x < 2 \\ 1, & \text{if } 2 \leq x < 4 \\ 0, & \text{if } 4 \leq x < 7 \\ \frac{1}{33}, & \text{if } 7 \leq x < 40 \end{cases}$$

$$c_{\text{dAARCUR}} : X^{(\text{dAARCUR})} \longrightarrow [0,\infty), \qquad x \mapsto \begin{cases} 0, & \text{if } -25 \leq x < -20 \\ \frac{1}{5}, & \text{if } -20 \leq x < -5 \\ \frac{1}{4}, & \text{if } -5 \leq x < -1 \\ 0, & \text{if } -1 \leq x < 1 \\ \frac{1}{4}, & \text{if } 1 \leq x < 5 \\ \frac{1}{5}, & \text{if } 5 \leq x < 20 \\ 0, & \text{if } 20 \leq x < 25 \end{cases}$$

In this connection it should be emphasised that the imprecision of the measured dREV values suggests the choice of minor distinguishability in an environment of 0 in order to avoid control actions that refer to stochastic error processes rather than to important state changes.

Table 3 specifies for each fuzzy set a corresponding value so that the fuzzy set represents the set of all values being indistinguishable from the considered value in the vague environment induced by the corresponding scaling function.

In this way, we can transform the rule base provided in Table 2 into the partial control mapping shown in Table 4.

The idea of fuzzy control as interpolation in vague environments was further developed by S. Kovács and L.T. Kóczy. First of all they propose to use an approximate scaling function in case the prerequisites of Theorem 2 are not satisfied [17]. In addition, the scaling functions are directly used to interpolate the output value. The idea is to choose an output value whose (scaled) distance to a reference output point coincides with the (scaled) distance of the actual input to a reference input point.

In [19] a method is discussed how the number of rules of a fuzzy controller can be reduced by using an approximate scaling function. [18] describes a successful application of fuzzy control on the basis of vague environments to the control of an automated guided vehicle.

				gREV			
	φ_0	-40	-6	-3	0	3	6	40
	-70	20	15	15	10	10	5	5
	-50	20	15	10	10	10	5	0
	-30	15	10	5	5	5	0	0
dREV	0	5	5	0	0	0	-10	-5
	30	0	0	0	-5	-5	-10	-15
	50	0	-5	-5	-10	-15	-15	-20
	70	-5	-5	-10	-15	-15	-15	-15

Table 4: The partial mapping φ_0 for idle speed control

4 A Regression Technique for Fuzzy Control

In the previous section fuzzy control was explained on the basis of scaling functions characterising vague environments. Instead of the specification of suitable fuzzy sets and control rules in the context of vague environments a partial control mapping and scaling functions have to be determined.

There are three principal approaches to the design of a fuzzy controller.

- A control expert is able to formulate appropriate control rules and fuzzy sets based on his experience and his knowledge about the process. This is usually a very tedious task and at least a fine tuning of the fuzzy sets (by hand) cannot be avoided.
- An experimental environment (the real process or a simulation of it) is available. Good control actions can be distinguished from bad ones. Then a neuro-fuzzy model like the one described in [27] or techniques based on evolutionary algorithms as in [6, 8] can be used to automatically generate a fuzzy controller.
- Data from observations of a control expert who is able to handle the process are available. In this case neuro-fuzzy methods (for an overview see [28]) as well as evolutionary algorithms (see for instance [30]) are applicable.

In this section we develop a regression technique that is tailored for fuzzy control in the spirit of vague environments for the above mentioned third case, when observation data are available.

We consider a simple Sugeno type fuzzy controller here which can be interpreted as a slightly modified version of Mamdani's model. The rules are of the form

$$R: \text{ if } \xi_1 \text{ is } A_1^{(R)} \text{ and } \ldots \text{ and } \xi_n \text{ is } A_n^{(R)} \text{ then } \eta \text{ is } b_R$$

where each linguistic term $A_1^{(R)}, \ldots, A_n^{(R)}$ is associated with a fuzzy set $\mu_1^{(R)}, \ldots, \mu_n^{(R)}$ on $X_1, \ldots, X_n$, respectively. b_R is a (crisp) output value assigned to the rule R.

For the aggregation of the premises of the rules we only assume that it is carried out by the t-norm $\odot$ (a commutative, associative, monotone increasing binary operation operation on the unit interval having 1 as unit, see for instance [20]), for instance the minimum or the product. The output value for the input tuple $(x_1, \ldots, x_n)$ is defined by the formula

$$f(x_1, \ldots, x_n) = \frac{\sum_R \left(\bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu) \right) \cdot b_R}{\sum_R \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu)} \tag{15}$$

where the sum in the nominator and denominator is defined for a finite set $\mathcal{R}$ of rules $R \in \mathcal{R}$.

Let us assume we have a set

$$D = \{(x_1^{(1)}, \ldots, x_n^{(1)}, y^{(1)}), \ldots, (x_1^{(s)}, \ldots, x_n^{(s)}, y^{(s)})\}$$

of sample data where the output $y^{(i)}$ is assigned to the input $(x_1^{(i)}, \ldots, x_n^{(i)})$. Let us for the moment consider the fuzzy sets μ_ν^R as given. For any set of parameters b_R $(R \in \mathcal{R})$ we can compute the quadratic error that is caused by the fuzzy controller with respect to the data set:

$$E = \sum_{\ell=1}^{s} \left(f(x_1^{(\ell)}, \ldots, x_n^{(\ell)}) - y^{(\ell)} \right)^2. \tag{16}$$

In order to minimise E, we have to choose the parameters b_R appropriately. To determine the b_R we take the partial derivatives of E with respect to the b_{R_0} $(R_0 \in \mathcal{R})$ and require them to be zero:

$$\frac{\partial E}{\partial b_{R_0}} = 0 \qquad (R_0 \in \mathcal{R}). \tag{17}$$

We obtain

$$\frac{\partial E}{\partial b_{R_0}} = \sum_{\ell=1}^{s} 2 \cdot \left(f(x_1^{(\ell)}, \ldots, x_n^{(\ell)}) - y^{(\ell)} \right) \cdot \frac{\partial f(x_1^{(\ell)}, \ldots, x_n^{(\ell)})}{\partial b_{R_0}}.$$

Taking the partial derivative of (15) and replacing f by (15) we get

$$\begin{aligned}
\frac{\partial E}{\partial b_{R_0}} &= \sum_{\ell=1}^{s} 2 \cdot \left(f(x_1^{(\ell)}, \ldots, x_n^{(\ell)}) - y^{(\ell)} \right) \cdot \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\sum_R \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)})} \\
&= 2 \cdot \left(\sum_{\ell=1}^{s} \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\left(\sum_R \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)}) \right)^2} \cdot \sum_R b_R \left(\bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)}) \right) \right. \\
&\qquad \left. - \sum_{\ell=1}^{s} \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\sum_R \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)})} \cdot y^{(\ell)} \right) \\
&= 2 \cdot \left(\sum_R b_R \sum_{\ell=1}^{s} \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\left(\sum_{R'} \bigodot_{\nu=1}^{n} \mu_\nu^{(R')}(x_\nu^{(\ell)}) \right)^2} \cdot \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)}) \right. \\
&\qquad \left. - \sum_{\ell=1}^{s} \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\sum_R \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)})} \cdot y^{(\ell)} \right) \qquad (18) \\
&= 0.
\end{aligned}$$

Thus (16) provides the following system of linear equations from which we can compute the b_{R_0} $(R_0 \in \mathcal{R})$:

$$\sum_R b_R \sum_{\ell=1}^{s} \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\left(\sum_{R'} \bigodot_{\nu=1}^{n} \mu_\nu^{(R')}(x_\nu^{(\ell)})\right)^2} \cdot \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)}) \;=\; \sum_{\ell=1}^{n} \frac{\bigodot_{\nu=1}^{n} \mu_\nu^{(R_0)}(x_\nu^{(\ell)})}{\sum_R \bigodot_{\nu=1}^{n} \mu_\nu^{(R)}(x_\nu^{(\ell)})} \cdot y^{(\ell)}. \tag{19}$$

A similar least square method was proposed in [9] for the identification of non-linear dynamic systems. We want to go a step further and tune the fuzzy sets in addition to the parameters b_R. Of course, this will not be possible by linear regression, since we would immediately run into a non-linear optimisation problem. Nevertheless, remembering the concept of fuzzy control based on vague environments, we can provide a good heuristics for tuning the fuzzy sets. We assume that the fuzzy sets on the domains $X_1, \ldots, X_n$ represent vague values, i.e. each of them corresponds to a fuzzy sets of values that are indistinguishable from a considered value. The vague environments are characterised by scaling functions that are constant between two neighbouring values which are used to build the fuzzy sets that we obtain fuzzy partitions of the type already illustrated in Figure 6.

The idea is the following. We start with a constant scaling factor leading to a homogeneous fuzzy partition. Then we compute the corresponding b_{R_0} from (19). In an area where we have a large error, it is necessary to specify the function f in more detail, i.e., we need more and narrower fuzzy sets. Therefore, we have to choose a larger scaling factor for such an area. Of course, we have to accept smaller scaling factors for areas with a small error. Otherwise we would increase the number of fuzzy sets what we do not want to consider here. This concept reflects the philosophy of large scaling factors for areas where the process (control function) is very sensitive to small changes and small scaling factors where the process (control function) does not change drastically with a variation of the parameters. Figure 8 illustrates the idea of contracting and moving fuzzy sets in an area with a large error.

In order to determine the change of the scaling factors we compute for each domain X_ν and each area the error of the regression function. By an area we mean the interval between two neighbouring points that induce the corresponding fuzzy sets appearing in the rules.

$$\mathrm{error}(\mathrm{area}_i) \;=\; \sum_{(x_1^{(\ell)},\ldots,x_n^{(\ell)}):x_\nu^{(\ell)} \in \mathrm{area}_i} \left(y^{(\ell)} - f(x_1^{(\ell)}, \ldots, x_n^{(\ell)})\right)^2 \tag{20}$$

area_i is the interval between the $(i-1)$th and ith point in the domain X_ν that induce the fuzzy sets appearing in the rules (compare Figure 9).

Now the area_i should be contracted, resulting in a larger scaling factor, when the error is relatively high, whereas it can be stretched when the error is small. If L_i^{old} is the length of area_i then we define the new (relative) length of the ith area by

$$L_i^{\mathrm{rel}} \;=\; \frac{L_i^{\mathrm{old}}}{\mathrm{const} + \mathrm{error}(\mathrm{area}_i)}$$

where const is a positive constant that first of all avoids devision by zero when the error is zero for a certain area. const also determines how strong the contraction or stretching of the corresponding area is depending on the error. If const is small in comparison to the error, this will result in drastic changes whereas a large constant allows only very small variations.

Finally, it is necessary to normalise the relative length of each area so that the over all length is the same as before, i.e., the length of the interval X_ν.

$$L_i \;=\; L_i^{\mathrm{rel}} \cdot \frac{\sum_j L_i^{\mathrm{old}}}{\sum_j L_j^{\mathrm{rel}}}. \tag{21}$$

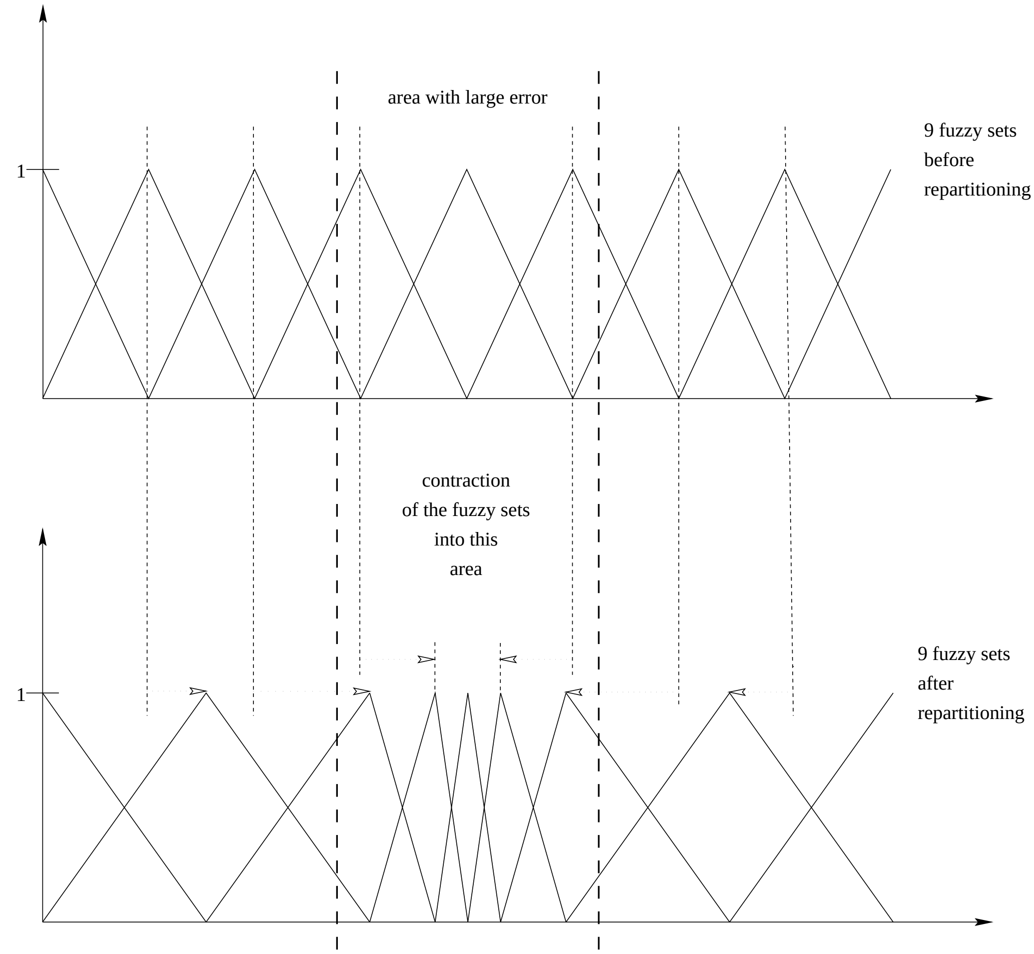

Figure 8: Contracting and moving fuzzy sets in an area with a large error

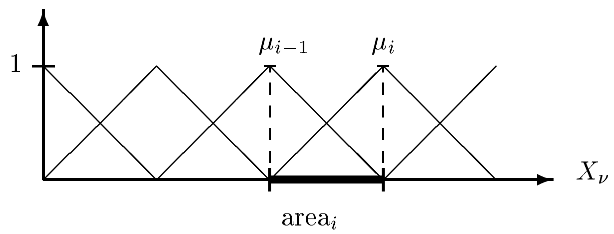

Figure 9: An area with a constant scaling factor

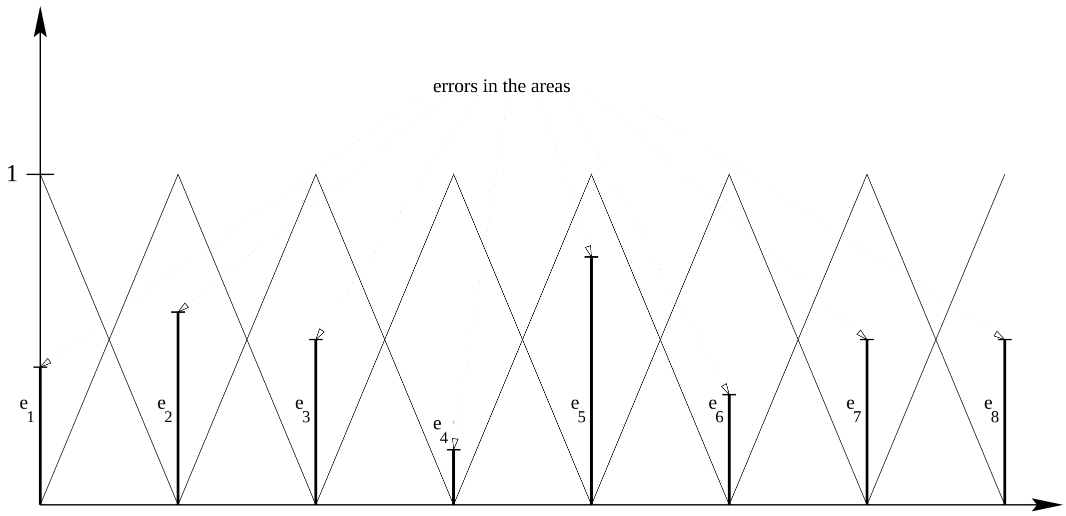

Figure 10: A fuzzy partition and errors for the corresponding areas

Figure 10 illustrates a fuzzy partition and indicates the magnitude of the error of the regression function for each area. The result of computing the updated length of each interval by formula (21) and building the new fuzzy partition by taking the induced fuzzy sets of the points at the boundaries of the areas is shown in Figure 11. Note that the corresponding scaling function is piecewise constant taking the value $1/L_i$ on the interval area$_i$.

Let us illustrate the approximation technique we have developed by two examples. In both cases we have chosen the product for the t-norm $\odot$. The first example is the piecewise linear function shown on the left hand side of Figure 12 from which we take 17 equidistant sample points. We start with the homogeneous fuzzy partition at the bottom of the left hand side of Figure 12. Then the values b_R in the conclusions in the rules are determined as the solution of the system of linear equations (19). After that we compute the error of the regression function for each subinterval (area) and obtain a new fuzzy partition by stretching or contracting the subintervals depending on the magnitude of the error. Then we determine new values b_R, obtain new errors for the (new) areas. This procedure is iterated until the error is sufficiently small or no significant improvements of the overall error were achieved during the last iteration steps. Finally, we end up with the fuzzy partition shown at the bottom of the right hand side of Figure 12 and the (nearly) perfect approximation (see the corresponding graph in the figure). Note that in the ranges where the function remains linear over a long interval a few wide fuzzy sets (small scaling factors) are chosen whereas in ranges where the function varies, more and narrower fuzzy sets (greater scaling factors) appear.

The second example is the two-dimensional function $\sin(x) \cdot \cos(y)$ in the range $[0, 2\pi]^2$ from which we take 1849 equidistant sample data. We start with a homogeneous fuzzy partition with eight fuzzy sets on each of the two domains, resulting in a rule base of $8 \cdot 8 = 64$ rules. Already after two iterations we obtain the modified fuzzy partitions illustrated in Figure 14 with the resulting regression function in Figure 15 which reduced the error to 10% of the error of the first regression function with homogeneous fuzzy partitions.

It should be emphasised that we do not aim at developing a new approximation technique that is superior to standard methods. The main point is that we are interested in rules that describe the function approximately and are understandable and interpretable for staff without strong mathematical training or background. Thus, the aim is knowledge extraction instead of best approximation.

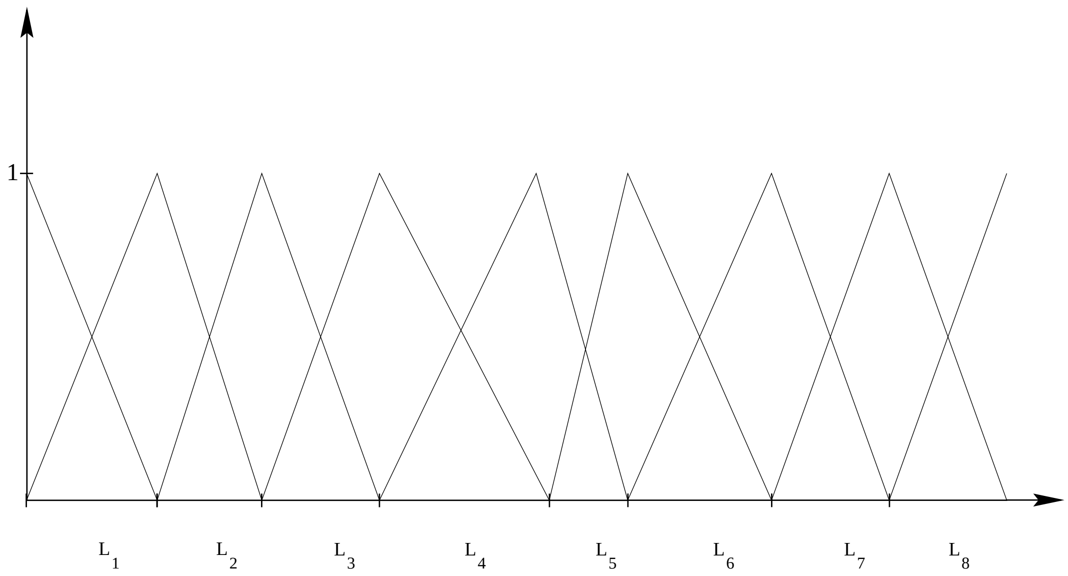

Figure 11: The new fuzzy partition

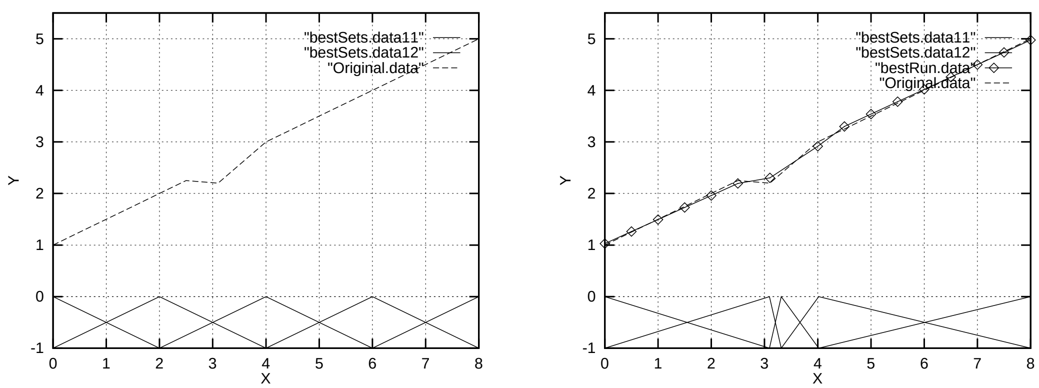

Figure 12: Approximation of a one-dimensional function

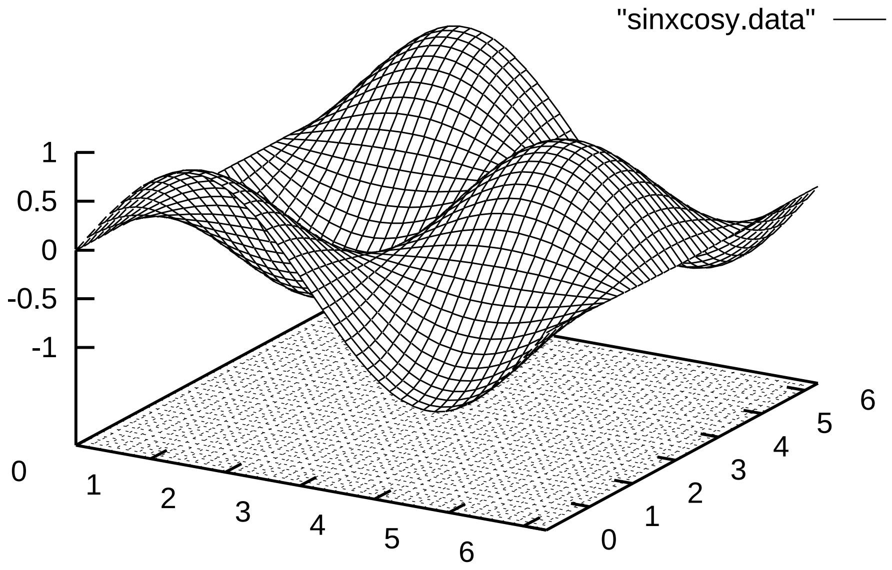

Figure 13: The function $\sin(x) \cdot \cos(y)$

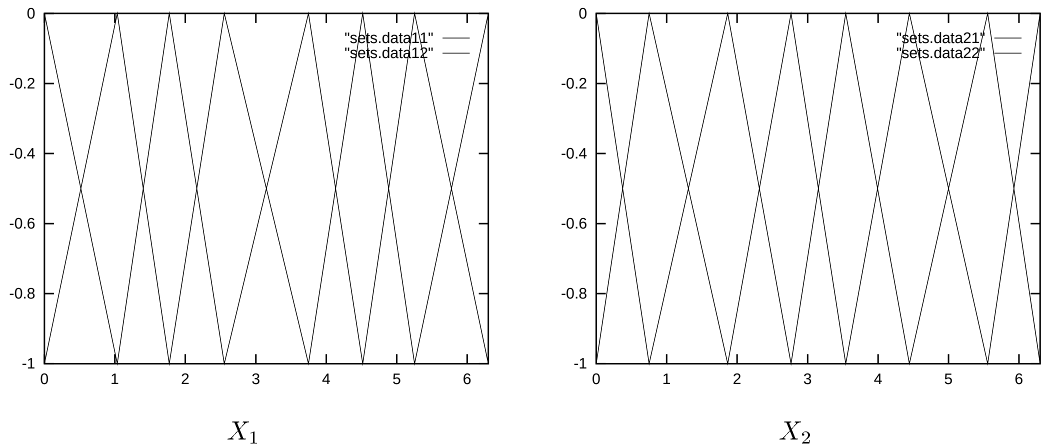

Figure 14: Modified fuzzy partitions after two iteration steps

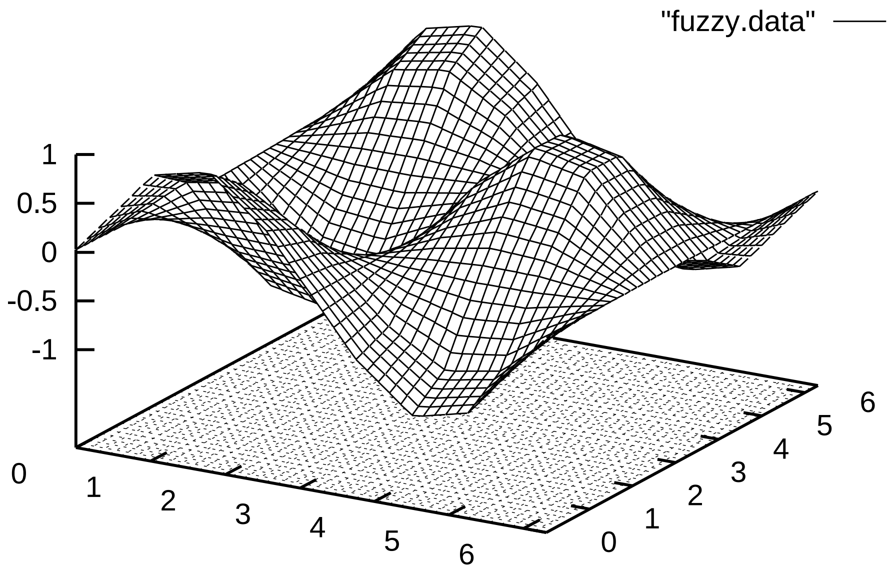

Figure 15: The resulting regression function

5 Conclusions

The semantics for fuzzy sets we have provided in this paper leads to a restriction of the possible parameter choices (e.g. arbitrary fuzzy partitions are not allowed) and a better understanding of the underlying assumptions (e.g. the indistinguishabilities in the different domains are considered independent). This can definitely simplify the design process of a fuzzy controller. We have also introduced a regression technique based on our considerations that enables to learn a fuzzy controller from data. Other techniques that are also related to this idea are based on fuzzy clustering (see for instance [15], for an overview see [7]). Fuzzy clusters are also fuzzy sets induced by a point, the so-called prototype or cluster centre. The membership degree is a function decreasing with increasing distance from the cluster centre.

Even if the semantics of vague environments may not always seem to be appropriate for a certain application, the concept of indistinguishability is inherent in fuzzy sets an cannot be avoided when operating with fuzzy sets [11].

Another advantage of vague environments is that the transformations induced by the scaling functions can be exploited in order to get more efficient computation schemes for fuzzy controllers [14].

References

[1] R.E. Bellman, R. Kalaba, L.A. Zadeh, Abstraction and Pattern Classification. Journ. Math. Anal. & Appl. 13 (1966), 1-7.

[2] R.E. Bellman, L.A. Zadeh, Decision Making in a Fuzzy Environment. Management Sciences 17 (1970), 141-164.

[3] D. Dubois and H. Prade, Fuzzy Sets in Approximate Reasoning, Part 1: Inference with Possibility Distributions. Fuzzy Sets and Systems 40 (1991), 143-202.

[4] D. Dubois, H. Prade, Similarity-Based Approximate Reasoning. In: J.M. Zurada, R.J. Marks II, C.J. Robinson (eds.), Computational Intelligence Imitating Life. IEEE Press, New York (1994), 69-80.

[5] J. Gebhardt, R. Kruse, The Context Model – An Integrating View of Vagueness and Uncertainty. Intern. Journ. of Approximate Reasoning 9 (1993), 283-314.

[6] J. Hopf, F. Klawonn, Learning the Rule Base of a Fuzzy Controller by a Genetic Algorithm. In: [21], 63-74.

[7] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis. Kluwer (to appear).

[8] J. Kinzel, F. Klawonn, R. Kruse, Modifications of Genetic Algorithms for Designing and Optimizing Fuzzy Controllers. Proc. IEEE Conference on Evolutionary Computation, IEEE, Orlando (1994), 28-33.

[9] V. Kecman, B.-M. Pfeiffer, Exploiting the Structural Equivalence of Learning Fuzzy Systems and Radial Basis Function Neural Networks. Proc. Second European Congress on Intelligent Techniques and Soft Computing (EUFIT'94), Aachen (1994), 58-66.

[10] F. Klawonn, Fuzzy Sets and Vague Environments. Fuzzy Sets and Systems 66 (1994), 207-221.

[11] F. Klawonn, J.L. Castro, Similarity in Fuzzy Reasoning. Mathware and Soft Computing 2 (1995), 197-228

[12] F. Klawonn, J. Gebhardt, R. Kruse, Fuzzy Control on the Basis of Equality Relations – with an Example from Idle Speed Control. IEEE Transactions on Fuzzy Systems 3 (1995), 336-350.

[13] F. Klawonn, R. Kruse, A Łukasiewicz Logic Based Prolog. Mathware and Soft Computing 1 (1994), 5-29.

[14] F. Klawonn, R. Kruse, Fuzzy Partitions and Transformations. Proc. 3rd IEEE Conference on Fuzzy Systems, IEEE, Orlando (1994), 1269-1273.

[15] F. Klawonn, R. Kruse, Constructing a Fuzzy Controller from Data. Fuzzy Sets and Systems 85 (1997), 177-193.

[16] F. Klawonn, V. Novák, The Relation between Inference and Interpolation in the Framework of Fuzzy Systems. Fuzzy Sets and Systems 81 (1996), 331-354.

[17] S. Kovács, New Aspects of Interpolative Reasoning. In: Proc. 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'96), Granada (1996), 477-482.

[18] S. Kovács, L.T. Kóczy, Application of the Approximate Fuzzy Reasoning Based on Interpolation in the Vague Environment of the Fuzzy Rulebase in the Fuzzy Logic Controlled Path Tracking Strategy of Differential Steered AGVs. In: B. Reusch (ed.), Computational Intelligence: Theory and Applications. Springer, Berlin (1997), 456-467.

[19] S. Kovács, L.T. Kóczy, Approximate Fuzzy Reasoning Based on Interpolation in the Vague Environment of the Fuzzy Rulebase as a Practical Alternative of the Classical CRI. In: Proc. 7th Int. Fuzzy Systems Association World Congress (IFSA'97) Vol. II, Prague (1997), 144-149.

[20] R. Kruse, J. Gebhardt, F. Klawonn, Foundations of Fuzzy Systems. Wiley, Chichester (1994).

[21] R. Kruse, J. Gebhardt, R. Palm (eds.), Fuzzy Systems in Computer Science. Vieweg, Braunschweig (1994).

[22] R. Kruse, E. Schwecke, and J. Heinsohn, Uncertainty and Vagueness in Knowledge Based Systems: Numerical Methods. Springer, Berlin (1991).

[23] C.C. Lee, Fuzzy Logic in Control Systems: Fuzzy Logic Controller, Part II. IEEE Trans. Systems, Man, Cybernetics 20 (1990), 419-435.

[24] E.H. Mamdani, Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Systems. IEEE Trans. Comput. 26 (1977), 1182-1191.

[25] R.E. Moore, Interval Analysis. Prentice Hall, Englewood Cliffs, NJ (1966).

[26] D. Mundici, Ulam Games, Łukasiewicz Logic, and AF $C^{\star}$-Algebras. Fundamenta Informaticae 18 (1993), 151-161.

[27] D. Nauck, Building Neural Fuzzy Controllers with NEFCON-I. In: [21], 141-151.

[28] D. Nauck, F. Klawonn, R. Kruse, Foundations of Neuro-Fuzzy Systems. Wiley, Chichester (1997).

[29] S.M. Ulam, Adventures of a Mathematician. Scribner's, New York (1976).

[30] H. Surmann, A. Kanstein, K. Goser, Self–Organizing and Genetic Algorithms for an Automatic Design of Fuzzy Control and Decision Systems. Proc. First European Congress on Intelligent Techniques and Soft Computing (EUFIT'93), Aachen (1993), 1097-1104.

[31] L.A. Zadeh, Fuzzy Sets. Information and Control 8 (1965), 338-353.

[32] L.A. Zadeh, Similarity Relations and Fuzzy Orderings. Information Sciences 3 (1971), 177-200.

[33] L.A. Zadeh, Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Trans. Systems, Man, and Cybernetics 3 (1973), 28-44.

[34] L.A. Zadeh, The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Parts 1,2,3. Information Sciences 8: 199-249, 8: 301-357, 9: 43-80, (1975).

[35] L.A. Zadeh, Fuzzy Sets as a Basis for a Theory of Possibility. Fuzzy Sets and Systems 1 (1978), 3-28.