6. OPTIMAL STRATEGY OF SWITCHING REASONING METHODS IN FUZZY CONTROL

Michael H. Smith[1] and Vladik Kreinovich[2]

[1] *Computer Science Division, University of California, Berkeley, CA 95720*
email mhs@robotics.eecs.berkeley.edu

[2] *Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968*
email vladik@cs.ep.utexas.edu

Abstract. *Fuzzy control is a methodology that transforms control rules (described by an expert in words of a natural language) into a precise control strategy. There exist several versions of this transformation. The main difference between these versions is in how they interpret logical connectives "and" and "or", i.e., in other words, what* reasoning method *a version uses. Which of these versions should we choose? It turns out that on different stages of control, different reasoning methods lead to better control results. Therefore, a natural idea is not to fix a reasoning method once and forever, but to* switch *between different reasoning methods. In this chapter, we describe the optimal choice of the reasoning methods between which we switch.*

6.1 Why switch?

Before we describe why switching helps, let us briefly recall what a fuzzy control is.

6.1.1 Fuzzy control methodology: a brief intro. In the situations when we do not have the complete knowledge of the plant, we often have the experience of human operators who successfully control this plant. We would like to make an automated controller that uses their experience. With this goal in mind, an ideal situation is when an operator can describe his control strategy in precise mathematical terms. However, most frequently, the operators cannot do that (can you describe how exactly you drive your car?). Instead, they explain their control in terms of *rules* formulated in natural language (like "if the velocity is high, and the obstacle is close, break immediately"). *Fuzzy control* is a methodology that translates these natural-language rules into an automated control strategy. This methodology was first outlined by L. Zadeh [5] and experimentally tested by E. Mamdani [17] in the framework of *fuzzy* set theory [25] (hence the name). For many practical systems, this approach works fine. For the current state of fuzzy control the reader is referred to the surveys [24], [16], [1].

Specifically, the rules that we start with are usually of the following type:

$$\text{if } x_1 \text{ is } A_1^j \text{ and } x_2 \text{ is } A_2^j \text{ and... and } x_n \text{ is } A_n^j, \text{ then } u \text{ is } B^j$$

where x_i are parameters that characterize the plant, u is the control, and A_i^j, B^j are the terms of natural language that are used in describing the $j-$th rule (e.g., "small", "medium", etc).

The value u is an appropriate value of the control if and only if at least one of these rules is applicable. Therefore, if we use the standard mathematical notations & for "and", $\vee$ for "or", and $\equiv$ for "if and only if", then the property "u is an appropriate control" (which we will denote by $C(u)$) can be described by the following informal "formula":

$$C(u) \equiv \; (A_1^1(x_1) \,\&\, A_2^1(x_2) \,\&\, ... \,\&\, A_n^1(x_n) \,\&\, B^1(u)) \vee$$

$$(A_1^2(x_1) \,\&\, A_2^2(x_2) \,\&\, ... \,\&\, A_n^2(x_n) \,\&\, B^2(u)) \vee$$

$$...$$

$$(A_1^K(x_1) \,\&\, A_2^K(x_2) \,\&\, ... \,\&\, A_n^K(x_n) \,\&\, B^K(u))$$

Terms of natural language are described as membership functions. In other words, we describe $A_i^j(x)$ as $\mu j, i(x)$, the degree of belief that a given value x satisfies the property A_i^j. Similarly, $B^j(u)$ is represented as $\mu_j(u)$. Logical connectives & and $\vee$ are interpreted as some operations $f_\vee$ and $f_\&$ with degrees of belief (e.g., $f_\vee = \max$ and $f_\& = \min$). After these interpretations, we can form the membership function for control: $\mu_C(u) = f_\vee(p_1, ..., p_K)$, where

$$p_j = f_\&(\mu_{j,1}(x_1), \mu_{j,2}(x_2), ..., \mu_{j,n}(x_n), \mu_j(u))), \qquad j = 1, ..., K.$$

We need an automated control, so we must end up with a single value $\bar{u}$ of the control that will actually be applied. An operation that transforms a membership function into a single value is called a *defuzzification*. Therefore, to complete the fuzzy control methodology, we must apply some defuzzification operator D to the membership function $\mu_C(u)$ and thus obtain the desired value $\bar{u} = f_C(\vec{x})$ of the control that corresponds to $\vec{x} = (x_1, ..., x_n)$. Usually, the *centroid defuzzification* is used, when

$$\bar{u} = \frac{\int u\mu_C(u)\,du}{\int \mu_C(u)\,du}.$$

6.1.2 A simple example: controlling a thermostat. The goal of a thermostat is to keep a temperature T equal to some fixed value T_0, or, in other words, to keep the difference $x = T - T_0$ equal to 0. To achieve this goal, one can switch the heater or the cooler on and off and control the degree of cooling or heating. What we actually control is the rate at which the temperature changes, i.e., in mathematical terms, a derivative $\dot{T}$ of temperature with respect to time. So if we apply the control u, the behavior of the thermostat will be determined by the equation $\dot{T} = u$. In order to automate this control we must come up with a function $u(x)$ that describes what control to apply if the temperature difference x is known.

In many cases, the exact dependency of the temperature on the control is not precisely known. Instead, we can use our experience, and formulate reasonable control rules:

- If the temperature T is close to T_0, i.e., if the difference $x = T - T_0$ is negligible, then no control is needed, i.e., u is also negligible.
- If the room is slightly overheated, i.e., if x is positive and small, we must cool it a little bit (i.e., $u = \dot{x}$ must be negative and small).
- If the room is slightly overcooled, then we need to heat the room a little bit. In other terms, if x is small negative, then u must be small positive.

So, we have the following rules:

- if x is negligible, then u must be negligible;
- if x is small positive, then u must be small negative;
- if x is small negative, then u must be small positive.

In this case, u is a reasonable control if either:

- the first rule is applicable (i.e., x is negligible) and u is negligible; or
- the second rule is applicable (i.e., x is small positive), and u must be small negative;
- or the third rule is applicable (i.e., x is small negative), and u must be small positive.

Summarizing, we can say that u is an appropriate choice for a control if and only if either x is negligible and u is negligible, or x is small positive and u is small negative, etc. If we use the denotations $C(u)$ for "u is an appropriate control", $N(x)$ for "x is negligible", SP for "small positive", and SN for "small negative", then we arrive at the following informal "formula":

$$C(u) \equiv (N(x)\&N(u)) \vee (SP(x)\&SN(u)) \vee (SN(x)\&SP(u)).$$

If we denote the corresponding membership functions by μ_N, μ_{SP}, and μ_{SN}, then the resulting membership function for control is equal to

$$\mu_C(u) = f_\vee(f_\&(\mu_N(x), \mu_N(u)), f_\&(\mu_{SP}(x), \mu_{SN}(u)), f_\&(\mu_{SN}(x), \mu_{SP}(u))).$$

6.1.3 It is reasonable to switch. There exist several versions of fuzzy control methodology. The main difference between these versions is in how they translate logical connectives "or" and "and", i.e., in other words, what *reasoning method* a version uses. Which of these versions should we choose? It turns out that on different stages of the plant, different reasoning methods lead to better control results (see, e.g., [21–23]). Therefore, a natural idea is not to fix a reasoning method once and forever, but to *switch* between different reasoning methods.

How to choose the methods between which we will switch? The goal of this Chapter is to give an answer to this question.

6.1.4 The contents of this Chapter. We will proceed as follows:

- The main criterion for choosing a set of reasoning methods is to achieve the best control possible. So, before we start the description of our problem, it is necessary to explain when a control is good. This will be done (first informally, then formally) in Section 6.2.
- Now that we know what our objective is, we must describe the possible choices, i.e., the possible reasoning methods. This description is given in Section 6.3.
- We are going to prove several results explaining what choice of a reasoning method leads to a better control. The proofs will be very general. However, for the readers' convenience, we will explain them on the example of a simple plant. This simple plant that will serve as a testbed for different versions of fuzzy control will be described in Section 6.4.
- The formulation of the problem in mathematical terms is now complete. In Section 6.5, we state the results, and in Section 6.6, the proofs of these results. In Section 6.7, we mention that these results have been experimentally confirmed (for details, see [21–23]). Our results are summarized in Section 6.8.
- The last (short) section contains open problems.

6.2 Criteria for switching: what do we expect from an ideal control?

6.2.1 What is an ideal control? In some cases, we have a well-defined control objective (e.g., minimizing fuel). But in most cases, engineers do not explain explicitly what exactly they mean by an *ideal* control. However, they often do not hesitate to say that one control is better than another one. What do they mean by that? Usually, they draw a graph that describes how an initial perturbation changes with time, and they say that a control is good if this perturbation quickly goes down to 0 and then stays there.

In other words, in a typical problem, an ideal control consists of two stages:

- On the *first stage*, the main objective is to make the difference $x = X - X_0$ between the actual state X of the plant and its ideal state X_0 go to 0 as fast as possible.

- After we have already achieved the objective of the first stage, and the difference is close to 0, then the *second stage* starts. On this second stage, the main objective is to keep this difference close to 0 at all times. We do not want this difference to oscillate wildly, we want the dependency $x(t)$ to be as smooth as possible.

This description enables us to formulate the objectives of each stage in precise mathematical terms.

6.2.2 First stage of the ideal control: main objective. We have already mentioned in Section 6.1 that for readers' convenience, we will illustrate our ideas on some simple plants. So, let us consider the case when the state of the plant is described by a single variable x, and we control the first time derivative $\dot{x}$. For this case, we arrive at the following definition:

Definition 6.1 Let a function $u(x)$ be given (this function will be called a *control strategy*). By a *trajectory* of the plant, we understand the solution of the differential equation $\dot{x} = u(x)$. Let's fix some positive number M (e.g., $M = 1000$). Assume also that a real number $\delta \neq 0$ is given. This number will be called an *initial perturbation.* A *relaxation time* $t(\delta)$ for the control $u(x)$ and the initial perturbation δ is defined as follows:

- we find a trajectory $x(t)$ of the plant with the initial condition $x(0) = \delta$, and
- take as $t(\delta)$, the first moment of time starting from which $|x(t)| \leq |x(0)|/M$ (i.e., for which this inequality is true for all $t \geq t(\delta)$).

Comment.

- For *linear control*, i.e., when $u(x) = -kx$ for some constant k, we have $x(t) = x(0)\exp(-kt)$ and therefore, the relaxation time t is easily determined by the equation $\exp(-kt) = 1/M$, i.e., $t = \ln(M/k)$. Thus defined relaxation time does not depend on δ. So, for control strategies that use linear control on the first stage, we can easily formulate the objective: to minimize relaxation time. The smaller the relaxation time, the closer our control to the ideal.
- In the *general case*, we would also like to minimize relaxation time. However, in general, we encounter the following problem: For *non-linear control* (and fuzzy control *is* non-linear) the relaxation time $t(\delta)$ depends on δ. If we pick a δ and minimize $t(\delta)$, then we get good relaxation for this particular δ, but possibly at the expense of not-so-ideal behavior for different values of the initial perturbation δ.

What to do? The problem that we encountered was due to the fact that we considered a simplified control situation, when we start to control a system only when it is already out of control. This may be too late. Usually, no matter how smart the control is, if a perturbation is large enough, the plant will never stabilize. For example, if the currents that go through an electronic system exceed a certain level, they will simply burn the electronic components. To avoid that, we usually control the plant from the very beginning, thus preventing the values of x from becoming too large. From this viewpoint, what matters is how fast we go down for *small* perturbations, when $\delta \approx 0$.

What does "small" mean in this definition? If for some value δ that we initially thought to be small, we do not get a good relaxation time, then we will try to keep the perturbations below that level. On the other hand, the smaller the interval that we want to keep the system in, the more complicated and costly this control becomes. So, we would not decrease the admissible level of perturbations unless we get a really big increase in relaxation time. In other words, we decrease this level (say, from δ_0 to $\delta_1 < \delta_0$) only if going from $t(\delta_0)$ to $t(\delta_1)$ means decreasing the relaxation time. As soon as $t(\delta_1) \approx t(\delta_0)$ for all $\delta_1 < \delta_0$, we can use δ_0 as a reasonable upper level for perturbations.

In mathematical terms, this condition means that $t(\delta_0)$ is close to the limit of $t(\delta)$ when $\delta \to 0$. So, the smaller this limit, the faster the system relaxes. Therefore, this limit can be viewed as a reasonable objective for the first stage of the control.

Definition 6.2 By a *relaxation time* T for a control $u(x)$, we mean the limit of $t(\delta)$ for $\delta \to 0$.

So, *the main objective of the first stage of control is to maximize relaxation time.*

LEMMA 6.1 *If the control strategy* $u(x)$ *is a smooth function of* x*, then the relaxation time equals to ln* $M/(-u'(0))$*, where* u' *denotes the derivative of* u*.*

Comment. So the bigger this derivative, the smaller the relaxation time. Therefore, our objective can be reformulated as follows: *to maximize* $u'(0)$*.*

6.2.3 Second stage of the ideal control: main objective. After we have made the difference x go close to 0, the second stage starts, on which $x(t)$ has to be kept as smooth as possible. What does *smooth* mean in mathematical terms? Usually, we say that a trajectory $x(t)$ is smooth at a given moment of time t_0 if the value of the time derivative $\dot{x}(t_0)$ is close to 0. We want to say that a trajectory is smooth if $\dot{x}(t)$ is close to 0 for all t.

In other words, if we are looking for a control that is the smoothest possible, then we must find the control strategy for which $\dot{x}(t) \approx 0$ for all t. There are infinitely many moments of time, so even if we restrict ourselves to control strategies that depend on finitely many parameters, we will have infinitely many equations to determine these parameters.

In other words, we will have an *over-determined* system. Such situations are well-known in data processing, where we often have to find parameters $p_1, ..., p_n$ from an over-determined system $f_i(p_1, ..., p_n) \approx q_i, 1 \le i \le N$. A well-known way to handle such situations is to use the *least squares method*, i.e., to find the values of p_j for which the "average" deviation between f_i and q_i is the smallest possible. To be more precise, we minimize the sum of the squares of the deviations, i.e., we are solving the following minimization problem:

$$\sum_{i=1}^{N} \left(f_i(p_1, ..., p_n) - q_i)^2 \right) \to \min_{p_1,\ldots,p_n} .$$

In our case, $f_i = \dot{x}(t)$ for different moments of time t, and $q_i = 0$. So, least squares method leads to the criterion $\sum(\dot{x}(t))^2 \to \min$. Since there are infinitely many moments of time, the sum turns into an integral, and the criterion for choosing a control into $J(x(t)) \to \min$, where $J(x(t)) = \int (\dot{x}(t))^2 \, dt$. This value J thus represents a degree to which a given trajectory $x(t)$ is non-smooth. So, we arrive at the following definition:

Definition 6.3 Assume that a control strategy $x(t)$ is given, and an initial perturbation δ is given. By a *non-smoothness* $I(\delta)$ of a resulting trajectory $x(t)$, we understand the value $J(x) = \int_0^\infty (\dot{x}(t))^2 \, dt$.

Comments.
1. The least squares method is not only heuristic, it has several reasonable justifications. So, instead of simply borrowing the known methodology from data processing (as we did), we can formulate reasonable conditions for a functional J (that describes non-smoothness), and thus deduce the above-described form of J without using analogies at all. This is done in [11].

2. What control to choose on the second stage? Similarly to relaxation time, we get different criteria for choosing a control if we use values of non-smoothness that correspond to different δ. And similarly to relaxation time, a reasonable solution to this problem is to choose a control strategy for which in the limit $\delta \to 0$, the non-smoothness takes the smallest possible value.

Mathematically, this solution is a little bit more difficult to implement than the solution for the first stage: Indeed, the relaxation time $t(\delta)$ has a well-defined non-zero limit when $\delta \to 0$, while non-smoothness simply tends to 0. Actually, for linear control, $I(\delta)$ tends to 0 as δ^2. To overcome this difficulty and still get a meaningful limit of non-smoothness, we will divide $J(x)$ (and, correspondingly, $I(\delta)$) by δ^2 and only then, tend this ratio $\tilde{J}(x(t)) = \tilde{I}(\delta)$ to a limit. This division does not change the relationship between the functional and smoothness: indeed, if for some δ, a trajectory $x_1(t)$ is smoother than a trajectory $x_2(t)$ in the sense that $J(x_1(t)) < J(x_2(t))$, then, after dividing both sides by δ^2, we will get $\tilde{J}(x_1(t)) < \tilde{J}(x_2(t))$. So, a trajectory $x(t)$ for which $\tilde{J}(x)$ is smaller, is thus smoother.

As a result, we arrive at the following definition.

Definition 6.4 By a *non-smoothness* I of a control $u(x)$, we mean the limit of $I(\delta)/\delta^2$ for $\delta \to 0$.

The main objective of the second stage of control is to minimize non-smoothness.

6.3 What are the possible reasoning methods?

6.3.1 General properties of $\vee-$ and $\&-$operations: commutativity and associativity. In order to apply fuzzy control methodology, we must assign a truth value (also called *degree of belief*, or *certainty value*) $t(A)$ to every uncertain statement A contained in the experts' rules. Then, we must define $\vee-$ and $\&-$operations $f_\vee(a,b)$ and $f_\&(a,b)$ in such a way that for generic statements A and B, $t(A \vee B)$ is close to $f_\vee(t(A),t(B))$, and $t(A\&B)$ is close to $f_\&(t(A),t(B))$. Let us first describe properties that are general to both $\vee-$ and $\&-$operations.

Statements $A\&B$ and $B\&A$ mean the same. Hence, $t(A\&B) = t(B\&A)$, and it is therefore reasonable to expect that $f_\&(t(A),t(B)) = f_\&(t(B),t(A))$ for all A and B. In other words, it is reasonable to demand that $f_\&(a,b) = f_\&(b,a)$ for all a and b, i.e., that $f_\&$ is a *commutative* operation. Similarly, it is reasonable to demand that $f_\vee$ is a commutative operation.

Statements $(A\&B)\&C$ and $A\&(B\&C)$ also mean the same thing: that all three statements A, B, and C are true. Therefore, it is reasonable to demand that the corresponding approximations $f_\&(f_\&(t(A),t(B)),t(C))$ and $f_\&(t(A),f_\&(t(B),t(C))$ coincide. In mathematical terms, it means that an $\&-$operation must be *associative*. Similarly, it is reasonable to demand that an $\vee-$operation is associative. To make our exposition complete, let us give a precise mathematical definition.

Definition 6.5 A function $f:[0,1]\times[0,1]\to[0,1]$ is called *commutative* if $f(a,b)=f(b,a)$ for all a and b. It is called *associative* if $f(f(a,b),c) = f(a,f(b,c))$ for all a, b, c.

Comment. If a function f is commutative and associative, then the result of applying f to several values $a,b,...,c$ does not depend on their order. So, we can use a simplified notation $f(a,b,...,c)$ for $f(a,f(b,...c)...))$.

6.3.2 What are the possible $\vee-$operations? One of the most frequently used methods of assigning a certainty value $t(A)$ to a statement A is as follows (see, e.g., [2], [3]; [7], IV.1.d; [10]): we take several (N) experts, and ask each of them whether he believes that a given statement A is true (for example, whether he believes that 0.3 is negligible). If $N(A)$ of them answer "yes", we take the ratio $t(A) = N(A)/N$ as a desired certainty value. In other words, we take $t(A) = |S(A)|/N$, where $S(A)$ is the set of all experts (out of the given N) who believe that A is true, and $|S|$ denotes the number of elements in a given set S. Here, $S(A \vee B) = S(A) \cup S(B)$, hence, $N(A \vee B) = |S(A \cup B)| \leq |S(A)| + |S(B)| = N(A) + N(B)$. If we divide both sides of this inequality by N, we can conclude that $t(A \vee B) \leq t(A) + t(B)$. Also, since $N(A) \leq N$, we get $t(A) \leq 1$, hence, $t(A \vee B) \leq \min(t(A) + t(B), 1)$.

On the other hand, since $S(A) \subseteq S(A) \cup S(B)$, we have $|S(A)| \leq |S(A \vee B)|$ and hence, $t(A) \leq t(A \vee B)$. Similarly, $t(B) \leq t(A \vee B)$. From these two inequalities, we can deduce that $\max(t(A), t(B)) \leq t(A \vee B)$. So, we arrive at the following definition:

Definition 6.6 By an $\vee-$*operation*, we will understand a commutative and associative function $f_\vee : [0,1] \times [0,1] \to [0,1]$ for which $\max(a, b) \leq f_\vee(a, b) \leq \min(a + b, 1)$ for all a and b.

Comment. Another possibility to estimate $t(A)$ is to interview a single expert and express his degree of confidence in terms of the so-called *subjective probabilities* [19]. For this method, similar inequalities can be extracted from the known properties of (subjective) probabilities.

6.3.3 What are the possible $\&-$operations? Similarly to $\vee$, we can conclude that $S(A\&B) = S(A) \cap S(B)$, so $N(A\&B) \leq N(A)$, $N(A\&B) \leq N(B)$, hence $N(A\&B) \leq \min(N(A), N(B))$ and $t(A\&B) \leq \min(t(A), t(B))$.

On the other hand, a person does not believe in $A\&B$ iff either he does not believe in A, or he does not believe in B. Therefore, the number $N(\neg(A\&B))$ of experts who do not believe in $A\&B$ cannot exceed the sum $N(\neg A) + N(\neg B)$. The number $N(\neg(A\&B))$ of experts who do not believe in $A\&B$ is equal to $N - N(A\&B)$, and similarly, $N(\neg A) = N - N(A)$ and $N(\neg B) = N - N(B)$. Therefore, the above-mentioned inequality turns into $N - N(A\&B) \leq N - N(A) + N - N(B)$, which leads to $N(A\&B) \geq N(A) + N(B) - N$ and hence, to $t(A\&B) \geq t(A) + t(B) - 1$. Since $t(A\&B) \geq 0$, we have $t(A\&B) \geq \max(0, t(A) + t(B) - 1)$. So, we arrive at the following definition:

Definition 6.7 By an $\&-$*operation*, we will understand a commutative and associative function $f_\& : [0,1] \times [0,1] \to [0,1]$ for which $\max(0, a + b - 1) \leq f_\&(a, b) \leq \min(a, b)$ for all a and b.

Comment. The same formulas hold if we determine $t(A)$ as a subjective probability.

6.3.4 Problems with $\&-$operations. The definition that we came up with for an $\vee-$operation was OK, but with $\&-$operations, we have two problems:

- In some situations, an $\&-$operation can be unusable for fuzzy control. For example, if $f_\&(a,b) = 0$ for some $a > 0, b > 0$, then for some $x, \dot{x}, ...$ the resulting membership function for a control $\mu_C(u)$ can be identically 0, and there is no way to extract a value of the control $\bar{u}$ from such a function. For such situations, it is necessary to further restrict the class of possible $\&-$operations.
- We based our definition of an $\&-$operation on the classical meaning of &, according to which if one of the statements A and B are false, then $A\&B$ is false. This is not always what we mean by "and" when we formulate the control rules. For example, a person can formulate a rule for choosing a future spouse as follows: "if someone is intelligent, and good-looking, and cooks well, and loves me, (and satisfies 24 more conditions), then that someone will be a perfect spouse for me." Here, "and" is used, but it does not mean that someone who does not satisfy one of these 28 conditions is immediately disqualified. Such a rule usually means that if only one of the conditions is satisfied, then there is some possibility that it will be applied; and the more conditions are satisfied, the bigger is that possibility. In other words, in some situations, the natural-language connective "and" behaves like a logical "or".

  "Behaves like" does not mean that it coincides exactly with the natural-language "or". Indeed, a natural-language statement "A or B" is absolutely true if only one of A or B is true, while this weird "or" is only to some extent true in this case.

  This distinction between the human "and" and the logical "and" is not limited to a few abnormal cases. It has been experimentally shown [26] that the human usage of "and" is often better described by operations different from the above-described "and"-operations (e.g., by an operation $\tilde{f}_\&(a,b) = p \cdot ab + (1-p)(a+b)/2$ that has non-zero values when $a = 0$ and $b > 0$). However, "often" does not mean "always". There are other cases (especially in engineering applications), when the "and" that we use in natural language behaves exactly like a logical "and". For example, if a car control includes a rule like "if an obstacle is close, and the speed is high, then either brake hard, or turn immediately", then this rule is certainly *not* applicable if the obstacle is not close.

  This phenomenon must be taken into consideration if we really want to use the rules that experts have formulated.

In the following two subsections, we will describe how these two problems can be solved.

6.3.5 Solution to the first problem: correlated &−operations. We have already mentioned that to solve the first problem (that $\mu_C(u)$ is identically 0 and hence, no fuzzy control is defined), we must restrict the class of possible &−operations. The forthcoming restriction will be based on the following idea. If belief in A and belief in B were independent events (in the usual statistical sense of the word "independent"), then we would have $t(A\&B) = t(A)t(B)$. In real life, beliefs are not independent. Indeed, if an expert has strong beliefs in several statements that later turn out to be true, then this means that he is really a good expert. Therefore, it is reasonable to expect that his degree of belief in other statements that are actually true will be bigger than the degree of belief of an average expert. If A and B are statements with $t(A) > 1/2$ and $t(B) > 1/2$, i.e., such that the majority of experts believe in A and in B, this means that there is a huge possibility that both A and B are actually true. A reasonable portion of the experts are *good experts*, i.e., experts whose predictions are almost often true. All of these good experts will believe in A and in B and therefore, all of them will believe in $A\&B$.

Let us give an (idealized) numerical example of this phenomenon. Suppose that, say, 60% of experts are good, and $t(A) = t(B) = 0.7$. This means that at least some of these good experts believe in A, and some believe in B. Since we assumed that the beliefs of good experts usually come out right, it means that A and B are actually true. Therefore, because of the same assumption about good experts, all good experts will believe in A, and all good experts will believe in B. Therefore, all of them will believe in $A\&B$. Hence, $t(A\&B) \geq 0.6 > t(A) \cdot t(B) = 0.49$.

In general, we have a mechanism that insures that there is, in statistical terms, a positive *correlation* between beliefs in A and B. In mathematical terms, the total number $N(A\&B)$ of experts who believe in $A\&B$ must be larger than the number $N_{ind}(A\&B) = Nt(A)t(B) = N(N(A)/N)(N(B)/N)$ that corresponds to the case when beliefs in A and B are uncorrelated random events. So we come to a conclusion that the following inequality sounds reasonable: $t(A\&B) \geq t(A)t(B)$. So, we arrive at the following definition:

Definition 6.8 An &−operation will be called *correlated* if $f_\&(a, b) \geq ab$ for all a, b.

Comment. In this case, we are guaranteed that if $a > 0$ and $b > 0$, then $f_\&(a, b) > 0$, i.e., we do avoid the problem in question.

6.3.6 Solution to the second problem: aggregation operations as &. The second problem states that when we use "and" in the natural language, it sometimes corresponds to a kind of a logical "or" (although in the majority of the cases, it still means a logical "and"). To solve this problem, let us recall how a function $f_\&$ can be determined experimentally (a similar method have been used to find an appropriate $\&-$operation for MYCIN [20], [4], and also in the general fuzzy context [9], [18], [26]). For a given a, b, we find pairs of statements A_i, B_i $(1 \le i \le K)$, for which $t(A_i) \approx a$ and $t(B_i) \approx b$, and for each pair, we ask the experts about the truth value of $A_i \& B_i$. As a result, we get K values $y_i = t(A_i \& B_i)$, $1 \le i \le K$. We want to define the desired value $y = f_\&(a, b)$ in such a way that this value will be "in the average" the closest possible to the actual values y_i. The same logic that we used to find the best description of smoothness leads us to the conclusion that this "closest possible in the average" is best represented by the least squares method, i.e., by choosing y for which

$$\sum_{i=1}^{K} (y - y_i)^2 \to \min_y .$$

This minimization problem can be solved easily, and the result is $y = (y_1 + ... + y_K)/K$. In other words, as $f_\&(a, b)$, we take an (arithmetic) average of the values $t(A_i, B_i)$ for all the cases when $t(A_i) \approx a$ and $t(B_i) \approx b$.

Let us use this idea to find the proper representation $\tilde{f}_\&(a, b)$ of a natural-language "and" (tilde is added to stress that the resulting operation is *not* a typical $\&-$operation and is therefore, different from $\&-$operations in the sense of Definition 6.7). When we average over all possible cases of using this "and", we must take into considerations that, according to our description of the second problem, there are two types of cases:

- In the majority of the cases, this "and" corresponds to a logical "and" and is thus represented by an $\&-$operation $f_\&(a, b)$ in the sense of Definition 6.7.
- In some cases, the natural-language "and" corresponds to a logical "or" and is, therefore, represented by some $\vee-$operation. We have also mentioned that this weird "or" is different from the regular "or". Therefore, it is reasonable to expect that the corresponding $\vee-$operation will be different from the operation $f_\vee$ used in the definition of fuzzy control. To express this difference, we will denote this new $\vee-$operation by $\tilde{f}_\vee$.

Let us denote the fraction of cases in which natural-language "and" is represented by a logical "and" by α. Then, the averaging over these two types of cases leads to the following formula for $\tilde{f}_\&$: $\tilde{f}_\&(a, b) = \alpha f_\&(a, b) + (1 - \alpha)\tilde{f}_\vee(a, b)$. So, to find an expression for the new operation, it is sufficient to describe $\tilde{f}_\vee(a, b)$, a "weird" operation that is related to a given $\&-$operation $f_\&(a, b)$.

This description can be obtained as follows. It is known that for every two sets S and S', $|S \cup S'| = |S| + |S'| - |S \cap S'|$. In particular, for $S = S(A)$ and $S' = S(B)$, we conclude that $N(A \vee B) = |S(A) \cup S(B)| = |S(A)| + |S(B)| - |S(A) \cap S(B)| = N(A) + N(B) - N(A\&B)$. If we divide both sides of this equality by N, we get $t(A \vee B) = t(A) + t(B) - t(A\&B)$. If we use $f_\&$ as an $\&-$operation, then we get $t(A\&B) \approx f_\&(t(A), t(B))$ and hence, $t(A \vee B) \approx a + b - f_\&(a, b)$, where we denoted $a = t(A)$ and $b = t(B)$. Therefore, it is natural to take the right-hand side of this inequality as a $\vee-$operation associated with a given $\&-$operation $f_\&$.

In other words, we define $\tilde{f}_\vee(a, b) = a + b - f_\&(a, b)$. If we substitute this expression into the formula for $\tilde{f}_\&$, we conclude that $\tilde{f}_\&(a, b) = \alpha f_\&(a, b) + (1-\alpha)(a+b-f_\&(a, b)) =$ (after a simple transformation) $= (2\alpha - 1)f_\&(a, b) + (1-\alpha)(a + b)$. This expression can be further simplified: Namely, if we denote $2\alpha - 1$ by p, then $\alpha = (1/2)(p + 1)$ hence, $1 - \alpha = (1/2)(1 - p)$, and $\tilde{f}_\&(a, b) = pf_\&(a, b) + (1/2)(1 - p)(a + b)$.

The resulting operation is an "aggregation" of "and" and "or", so it is usually called an *aggregation operation*. So, we arrive at the following definition:

Definition 6.9 Assume that a constant $p \in (0, 1)$ is given. By an *aggregation $\&-$operation*, we mean a function $\tilde{f}_\&(a, b) = pf_\&(a, b) + (1/2)(1-p)(a+b)$, where $f_\&$ is an $\&-$operation in the sense of Definition 6.7.

Comments.
1. For a list of known aggregation operations, see, e.g., [8], Section 3.1.2.3, or [27], Chapter 3. Experimental analysis (Czagola, 1988) has shown that the following two operations are the best in describing the human usage of "and": $\alpha \cdot ab + (1 - \alpha)(a + b - ab)$ and $\alpha \min(a, b) + (1-p)(a+b-\min(a, b))$ $(= \alpha \min(a, b) + (1-\alpha)\max(a, b))$. Both operations satisfy our Definition of an aggregation $\&-$operation. This fact can be viewed as an additional confirmation that the arguments that we used to justify this definition were indeed reasonable.

2. Unlike a normal fuzzy "and", aggregation "and"-operation is not associative. Therefore, we cannot automatically apply it to the case when we have three or more statements combined by "and": we need some further research. In view of that restriction, we will only use this operation for the case when each rule has no more than 2 conditions.

Usually in fuzzy control, the number of conditions coincides with the number of input variables. Therefore, our applications of aggregation operations to fuzzy control will be restricted to the plants whose states can be described by specifying the values of only two variables.

3. When we apply this aggregation operation to fuzzy control, we must also take into consideration the following difference:

- Aggregation "and" means that a rule can be fired even if not all of its conditions are satisfied. So, it must be used to combine the truth values of different conditions in a rule.
- On the other hand, the conclusion of the rule was combined with its condition also by an "and"-operation. For this combination, applying an aggregation "and" would mean that we are applying a rule even if none of its conditions are satisfied. This makes no sense. Therefore, for this particular combination, only a normal &−operation should be applied, and never an aggregation one.

As a result, if we use an aggregation operation $\tilde{f}_\&(a,b) = pf_\&(a,b) + (1/2)(1-p)(a+b)$ for "and", we arrive at the following expression for the membership function for control: $\mu_C(u) = f_\vee(p_1, ..., p_K)$, where $p_j = f_\&(c_j, \mu_j(u))$ and for each rule j, the expression for c_j depends on how many conditions this rule contains:

- If a rule contains only one condition $A_1^j(x_1)$, then $c_j = \mu_{j,1}(x_1)$.
- If a rule contains two conditions, then $c_j = \tilde{f}_\&(\mu_{j,1}(x_1), \mu_{j,2}(x_2))$.

6.4 Let's describe a simplified plant, on which different reasoning methods will be tested

6.4.1 Plant. Following Section 6.2, we will consider the simplest case when the state of the plant is described by a single variable x, and we control the first time derivative $\dot{x}$. To complete our description of the control problem, we must also describe:

- the experts' *rules*,
- the corresponding *membership functions*, and
- defuzzification.

6.4.2 Membership functions. For simplicity, we will consider the simplest (and most frequently used; see, e.g., [13–15]) membership functions, namely, triangular ones (as we will see from our proof, the result will not change if we use any other type of membership functions).

Definition 6.10 By a *triangular* membership function with a midpoint a and endpoints $a - \Delta_1$ and $a + \Delta_2$ we mean the following function $\mu(x)$:

- $\mu(x) = 0$ if $x < a - \Delta_1$ or $x > a + \Delta_2$;
- $\mu(x) = (x - (a - \Delta_1))/\Delta_1$ if $a - \Delta_1 \le x \le a$;
- $\mu(x) = 1 - (x - a)/\Delta_2$ if $a \le x \le a + \Delta_2$.

6.4.3 Rules. Fuzzy control can be viewed as a kind of extrapolation. In reality there exists some control $u(x, ...)$ that an expert actually applies. However, he cannot precisely explain, what function u he uses. So we ask him lots of questions, extract several rules, and form a fuzzy control from these rules.

We will restrict ourselves to the functions $u(x)$ that satisfy the following properties:

Definition 6.11 By an *actual control function* (or *control function*, for short), we mean a function $u(x)$ that satisfies the following three properties:

- $u(0) = 0$;
- $u(x)$ is monotonically decreasing for all x;
- $u(x)$ is smooth (differentiable).

Comment. These restrictions are prompted by common sense:

- If $x = 0$, this means that we are already in the desired state, and there is no need for any control, i.e., $u(0) = 0$.
- The more we deviate from the desired state $x = 0$, the faster we need to move back if we want the plant to be controllable. So, u is monotonically decreasing.
- We want the control to be smooth (at least on the second stage), so the function $u(x)$ that describes an expert's control, must be smooth.

Let's now describe the resulting rules formally.

Definition 6.12 Let's fix some $\Delta > 0$. For every integer j, by N_j, we will denote a triangular membership function with a midpoint $j\Delta$ and endpoints $(j-1)\Delta$ and $(j+1)\Delta$. We will call the corresponding fuzzy property N_0 *negligible* (N for short), N_1 *small positive* or SP, and N_{-1} *small negative*, or SN. Assume that a monotonically non-increasing function $u(x)$ is given, and that $u(0) = 0$. By *rules generated by* $u(x)$, we mean the set of following rules: "if $N_j(x)$, then $M_j(u)$" for all u, where M_j is a triangular membership function with a midpoint $u(j\Delta)$ and endpoints $u((j-1)\Delta)$ and $u((j+1)\Delta)$.

Comments.

1. In particular, if we start with a linear control $u = -kx$ (and linear control is the one that is most frequently used, see. e.g., [6]), then M_j resembles N_{-j} with the only difference being that instead of Δ, we use $k\Delta$. So, we can reformulate the corresponding rules as follows: if x is negligible, then u must be negligible; if x is small positive, then u must be small negative, etc. Here, we use Δ when we talk about x, and we use $k\Delta$ when we talk about u.

2. How to choose Δ? We have two phenomena to take into consideration:

- On one hand, the smaller Δ, the better the resulting rules represent the original expert's control. From this viewpoint, the smaller Δ, the better.
- On the the other hand, the smaller Δ, the more rules we will have and therefore, the more running time our control algorithm will require. So, we must not take Δ too small.

As a result, the following is the natural way to choose Δ:

- choose some reasonable value of Δ;
- if the resulting control is not good enough, decrease Δ;
- repeat this procedure until the further decrease does not lead to any improvement in the control quality.

So, the quality (i.e., relaxation time or non-smoothness) of the rule-based control for the chosen Δ will be close to the limit value of this quality when $\Delta \to 0$. Therefore, when choosing the best reasoning method, we must consider this limit quality as a choosing criterion. Let's formulate the relevant definitions.

Definition 6.13 Assume that the following are given:

- an actual control function $u(x)$;
- a defuzzification procedure.

For a given $\Delta > 0$, by a $\Delta-$*relaxation time*, we mean the relaxation time of a control strategy that is generated by an actual control function $u(x)$ for this Δ. By a *relaxation time, corresponding to an actual control function* $u(x)$, we mean the limit of $\Delta-$relaxation times when $\Delta \to 0$.

Definition 6.14 Assume that the following are given:

- an actual control function $u(x)$;
- a defuzzification procedure.

For a given $\Delta > 0$, by a $\Delta-$*non-smoothness*, we mean the non-smoothness of a control strategy that is generated by an actual control function $u(x)$ for this Δ. By a *non-smoothness, corresponding to an actual control function* $u(x)$, we mean the limit of $\Delta-$non-smoothness when $\Delta \to 0$.

6.4.4 Defuzzification. For simplicity of analysis, we will only use centroid defuzzification.

The formulation of the problem in mathematical terms is now complete.

6.5 Main results

6.5.1 First stage: minimizing relaxation time.

THEOREM 6.1 *Assume that an actual control function $u(x)$ is given. Then, among all possible $\vee-$ and $\&-$operations, the smallest relaxation time, corresponding to $u(x)$, occurs when we use $f_\vee(a,b) = \min(a+b,1)$ and $f_\&(a,b) = \min(a,b)$.*

If we allow aggregation $\&-$operations, then the result will be different:

THEOREM 6.1A *Assume that a constant $p \in (0,1)$ that defines aggregation operations is fixed. Assume that an actual control function $u(x)$ is given. Then among all possible $\vee-$operations, and among all possible $\&-$operations and aggregation $\&-$operations, the smallest relaxation time, corresponding to $u(x)$, occurs when we use $f_\vee(a,b) = \min(a+b,1)$ and $\tilde{f}_\&(a,b) = p\min(a,b) + (1-p)(a+b)/2$.*

Comments.
1. For our simplified plant, we only use rules with one condition, so, strictly speaking, applying an aggregation operation $\tilde{f}_\&(a,b) = pf_\&(a,b) + (1/2)(1-p)(a+b)$ leads to the same fuzzy control as applying the corresponding $\&-$operation $f_\&(a,b)$. In other words, for this particular simplified plant, this Theorem is a trivial corollary of Theorem 6.1. However, as we have already mentioned, our results, although formulated for this simplified plant only, are actually applicable to a much more general case, because we do not use specific features of this plant in our proofs. In view of this, we decided to add this result (and to supply it with a proof that is applicable even when we have two conditions in each rule).

2. Theorems 6.1 and 6.1A mean that *on the first stage, we must use one of the following two reasoning methods:*

- $f_\vee(a,b) = \min(a+b,1)$ and $f_\&(a,b) = \min(a,b)$;
- $f_\vee(a,b) = \min(a+b,1)$ and $\tilde{f}_\&(a,b) = p\min(a,b) + (1-p)(a+b)/2$.

Theorem 6.1A *does not* mean that the aggregation operation is always preferable on the first stage of fuzzy control, because, as we have already mentioned, "and" in the rules often means logical "and" (in the sense that if one of the condition is not satisfied, then the rule is not applicable), and therefore, cannot be represented by an aggregation operator. In general, the meaning of "and" may change from rule to rule and, even within a rule, from case to case. So, between min and an aggregation operation, there is no ready-made preference. As we will see later, our experimental results confirm this conclusion in the sense that for one and the same plant, in different moments of time, different choices lead to a better control.

6.5.2 Second stage: minimizing non-smoothness.

THEOREM 6.2 *Assume that an actual control function $u(x)$ is given. Then among all possible $\vee-$ and $\&-$operations, the smallest non-smoothness, corresponding to $u(x)$, occurs when we use $f_\vee(a,b)=\max(a,b)$ and $f_\&(a,b)=\max(a+b-1,0)$.*

Comments.
1. Unlike the first stage, the result does not change if we allow aggregation operations.

2. We have already mentioned that since we are using an $\&-$operation for which $f_\&(a,b)=0$ for some $a,b>0$, we may end up with a situation when the resulting function $\mu_C(u)$ is identically 0 and therefore, fuzzy control methodology is not applicable. For such a situation, we must restrict ourselves to correlated $\&-$operations. For these operations, we get the following result:

THEOREM 6.2A *Assume that an actual control function $u(x)$ is given. Then among all possible $\vee-$operations and all possible correlated $\&-$operations, the smallest non-smoothness, corresponding to $u(x)$, occurs when we use $f_\vee(a,b)=\max(a,b)$ and $f_\&(a,b)=ab$.*

Comment. Theorems 6.2 and 6.2A mean that *on the second stage, we must use one of the following reasoning methods:*
- $f_\vee(a,b)=\max(a,b)$ and $f_\&(a,b)=\max(a+b-1,0)$, or
- $f_\vee(a,b)=\max(a,b)$ and $f_\&(a,b)=ab$.

Similarly to the first stage, the choice between these two $\&-$operations depends on the circumstances: for one and the same plant, at different moments of time, different $\&-$operations lead to a better (in this case, smoother) control.

General comment. These results are in good accordance with the general optimization results for fuzzy control described in [11].

6.6 Main ideas of the proofs

Proof of the Lemma is simple, because for small δ the control is approximately linear: $u(x)\approx u'(0)x$.

Proof of Theorem 6.1. Let us first consider the case when $u(x)$ is a linear function i.e., when $u(x)=-kx$. In this case, instead of directly proving the statement of Theorem 6.1 (that the limit of $\Delta-$relaxation times is the biggest for the chosen reasoning method), we will prove that for every Δ, $\Delta-$relaxation time is the largest for this very pair of $\vee-$

and &−operations. The statement itself will then be easily obtained by turning to a limit $\Delta \to 0$.

So, let us consider the case when $u(x) = -kx$ for some $k > 0$. In view of the Lemma, we must compute the derivative $\bar{u}'(0) = \lim_{x\to 0}\,(\bar{u}(x) - \bar{u}(0))/x)$, where $\bar{u}(x)$ is the control strategy into which the described fuzzy control methodology translates our rules.

It is easy to show that $\bar{u}(0) = 0$. Hence, $\bar{u}'(0) = \lim \bar{u}(x)/x$. So, to find the desired derivative, we must estimate $\bar{u}(x)$ for small x. To get the limit, it is sufficient to consider only negative values $x \to 0$. Therefore, for simplicity of considerations, let us restrict ourselves to small negative values x (we could as well restrict ourselves to positive x, but we have chosen negative ones because for them the control is positive and therefore, slightly easier to handle).

In particular, we can always take all these x from an interval $[-\Delta/2, 0]$. For such x, only two of the membership functions N_j are different from 0: $N(x) = N_0(x) = 1 - |x|/\Delta$ and $SN(x) = N_{-1}(x) = |x|/\Delta$. Therefore, only two rules are fired for such x, namely, those that correspond to $N(u)$ and $SP(u)$.

We have assumed the centroid defuzzification rule, according to which $\bar{u}(x) = n(x)/d(x)$, where the numerator $n(x) = \int u\mu_C(u)\,du$ and the denominator is equal to $d(x) = \int \mu_C(u)\,du$. When $x = 0$, the only rule that is applicable is $N_0(x) \to N_0(u)$. Therefore, for this x, the above-given general expression for $\mu_C(u)$ turns into $\mu_C(x) = \mu_N(u)$ Indeed, from our definitions of &− and ∨−operations, we can deduce the following formulas:

- $f_\&(a, 0) = 0$ for an arbitrary a, so the rule whose condition is not satisfied leads to 0, and
- $f_\vee(a, 0) = 0$ for all a, so the rule that leads to 0, does not influence $\mu_C(u)$.

Therefore, for $x = 0$, the denominator $d(0)$ equals $\int \mu_N(u)\,du = k\Delta$ (this is the area of the triangle that is the graph of the membership function).

So, when $x \to 0$, then $d(x) \to d(0) = k\Delta$. Therefore, we can simplify the expression for the desired value $\bar{u}'(0)$: $\bar{u}'(0) = \lim u(x)/x = \lim(n(x)/d(x))/x = (k\Delta)^{-1}\lim(n(x)/x)$. Since $k\Delta$ is a constant that does not depend on the choice of a reasoning method (i.e., of ∨− and &−operations), the biggest value of $\bar{u}'(0)$ (and hence, the smallest relaxation time) is attained when the limit $\lim(n(x)/x)$ takes the smallest possible value. So, from now on, let's estimate this limit.

For small negative x, as we have already mentioned, only two rules are fired: $N(x) \to N(u)$ and $SN(x) \to SP(u)$. Therefore, the membership function for control takes the following form: $\mu_C(u) = f_\vee(p_1(u), p_2(u))$, where $p_1(u) = f_\&(\mu_N(x), \mu_N(u))$ and $p_2(u) = f_\&(\mu_{SN}(x), \mu_{SP}(u))$. The function $\mu_{SP}(u)$ is different from 0 only for $u > 0$. Therefore, for $u < 0$, we have $p_2(u) = 0$ and hence, $\mu_C(u) = p_1(u)$.

We are looking for the reasoning method, for which $\lim(n(x)/x)$ takes the largest possible value, where $n(x) = \int \mu_C(u)\,du$. Let's fix an arbitrary &−operation $f_\&$ and consider different functions $f_\vee$. If we use two different ∨−operations $f_\vee(a,b)$ and $g_\vee(a,b)$ for which $f_\vee(a,b) \le g_\vee(a,b)$ for all a, b, then, when we switch from $f_\vee$ to $g_\vee$, the values of $\mu_C(u)$ for $u < 0$ will be unaffected, but the values for $u > 0$ will increase. Therefore, the total value of the numerator integral $n(x) = \int \mu_C(u)\,du$ will increase after this change. So, if we change $f_\vee$ to a maximum possible function $\min(a+b, 1)$, we will increase this integral. Therefore, we will arrive at a new pair of functions, for which the new value of $\bar{u}$ is not smaller for small x, and, therefore, the derivative of $\bar{u}$ in 0 is not smaller.

Therefore, when looking for the best reasoning methods, it is sufficient to consider only the pairs of ∨− and &−operations in which $f_\vee(a,b) = \min(a+b, 1)$. In this case, we have $\mu_C(x) = p_1(u) + p_2(u) - p_{ab}(u)$, where $p_{ab}(u)$ is different from 0 only for $u \approx 0$, and corresponds to the values u for which we use the 1 part of the $\min(a+b, 1)$ formula. Therefore, $n(x)$ can be represented as the sum of the three integrals: $n(x) = n_1 + n_2 - n_{ab}$, where $n_1 = \int u p_1(u)\,du$, $n_2 = \int u p_2(u)\,du$, and $n_{ab} = \int p_{ab}(u)\,du$. Let's analyze these three components one by one.

- The function $p_1(u)$ is even (because $\mu_N(u)$ is even). It is well known that for an arbitrary even function f, the integral $\int u f(u)\; du$ equals 0. Therefore, $n_1 = 0$. So, this component does not influence the limit $\lim(n(x)/x)$ (and therefore, does influence the relaxation time).
- The difference $p_{ab}(u)$ is of size $\sim u \sim x$, and it is different from 0 on the area surrounding $u = 0$ that is also of size $\sim x$. Therefore, the corresponding integral n_{ab} will be of order x^3. Therefore, when $x \to 0$, we have $n_{ab}/x \sim x^2 \to 0$. This means that this component does not influence the limit $\lim(n(x)/x)$ either.

As a result, the desired limit is completely determined by the second component $p_2(u)$, i.e., $\lim(n(x)/x) = \lim(n_2(x)/x)$. Therefore, the relaxation time is the smallest when $\lim(n_2(x)/x)$ takes the biggest possible value. Now, $n_2 = \int u p_2(u)\,du$, where $p_2(u) = f_\&(\mu_{SN}(x), \mu_{SP}(u))$. The membership function $\mu_{SP}(u)$ is different from 0 only for positive u. Therefore, the function $p_2(u)$ is different from 0 only for positive u. So, the bigger $f_\&$, the bigger n_2. Therefore, the maximum is attained, when $f_\&$ attains its maximal possible value, i.e., $\min(a,b)$. For linear actual control functions, the statement of the theorem is thus proven.

The general case follows from the fact that the relaxation time is uniquely determined by the behavior of a system near $x = 0$. The smaller Δ we take, the closer $u(x)$ to a linear function on an interval $[-\Delta, \Delta]$ that determines the derivative of $\bar{u}(x)$, and, therefore, the closer the corresponding relaxation time to a relaxation time of a system that originated from the linear control. Since for each of these approximating systems, the resulting relaxation time is the smallest for a given pair of $\vee-$ and $\&-$operations, the same inequality will be true for the original system that these linear systems approximate. Q.E.D.

Proof of Theorem 6.1A is similar, with the only difference in the choice of the biggest possible $\&-$operation: indeed,

- Since $f_\&(a,b) \le \min(a,b) \le (a+b)/2$, every aggregation $\&-$operation $\tilde{f}_\&(a,b) = pf_\&(a,b) + (1/2)(1-p)(a+b)$ is bigger than the corresponding $\&-$operation $f_\&(a,b)$. So, when we are looking for the biggest possible $\&-$operation, it is sufficient to look among aggregation $\&-$operations.
- An aggregation operation $\tilde{f}_\&(a,b) = pf_\&(a,b) + (1/2)(1-p)(a+b)$ is bigger than another aggregation operation $\tilde{g}_\&(a,b) = pg_\&(a,b) + (1/2)(1-p)(a+b)$ iff $f_\&(a,b) \ge g_\&(a,b)$ for all a and b. Therefore, the aggregation operation is the biggest possible iff the corresponding $\&-$operation is the biggest possible, i.e., when we use $f_\&(a,b) = \min(a,b)$.

Proof of Theorems 6.2 and 6.2A. For a linear system $u(x) = -kx$, we have $x(t) = \delta\ exp(-kt)$, so $\dot{x}(t) = -k\delta\ exp(-kt)$, and the non-smoothness functional equals $I(\delta) = \delta^2 \int_0^\infty k^2\ exp(-2kt)\,dt = (k/2)\delta^2$. Therefore, $I = k/2$. For non-linear systems with a smooth control $u(x)$ we can similarly prove that $I = -1/2u'(0)$. Therefore, the problem of choosing a control with the smallest value of non-smoothness is equivalent to the problem of finding a control with the smallest value of $k = |u'(0)|$. This problem is directly opposite to the problem that we solved in Theorem 6.1, where our main goal was to maximize k.

Similar arguments show that the smallest value of k is attained, when we take the smallest possible function for $\vee$, and the smallest possible operation for $\&$. Q.E.D.

Comment. We have proved our results only for the simplified plant. However, as one can easily see from the proof, we did not use much of the details about this plant. What we mainly used was the inequalities between different $\&-$ and $\vee-$operations. In particular, our proofs do not use the triangular form of the membership function, they use only the fact that the membership functions are located on the intervals $[a-\Delta, a+\Delta]$.

Therefore, a similar proof can be applied in a much more general context. We did not formulate our results in this more general context because we did not want to cloud our results with lots of inevitable technical details.

6.7 Experimental confirmation

In this section, we will simply mention these results briefly (for details, the reader is referred to [21–23]). The experiments included the application of different reasoning methods to a classical testbed of control: inverted pendulum on a cart. It is worth noticing that this problem also has practical applications, e.g., in assuring that during the launch, a missile is vertical. Rules for balancing the inverted pendulum are well known in fuzzy control literature (see, e.g., [12]).

The state of an inverted pendulum is described by the values of two variables:
x_1 : an angle between the pendulum and the vertical axis;
x_2 : angular velocity of the pendulum.
Since we have only two variables, we can (in principle) apply aggregation operations.

Inverted pendulum was simulated by a computer program with a small time quantum Δt. At every moment of time $(0, \Delta t,\ 2\Delta t,\ ...)$, 80 different reasoning methods were tried (the list of possible methods included 8 methods that use an aggregation $\&-$operation). Each reasoning technique leads to a slightly different change in a trajectory. Then, a method was chosen, for which the resulting trajectory was the closest to the ideal one (described in 6.2).

To make results more convincing, no theoretical assumptions were used to screen any of the reasoning methods, and for each moment of time, all 80 methods were tried.

As a result, we got a clear division into two stages:

- On the first stage, at all moments of time (with a few exceptions), one of the following two reasoning methods turned out to be the best:
  - $f_\vee(a,b) = \min(a+b,1)$ and $f_\&(a,b) = \min(a,b)$;
  - $f_\vee(a,b) = \min(a+b,1)$ and $\tilde{f}_\&(a,b) = p\min(a,b) + (1-p)(a+b)/2$.

  Switching between these two reasoning methods occurred at seemingly random moments of time, and we could not trace any regularity in those switches.

  The only regularity that we noticed was that initially (and we always started with $x_1 \neq 0$ and $x_2 = 0$), min worked better than an aggregation operation (in other words, we needed min for the initial push).
  A possible reason is that if we use an aggregation operation, then in addition to the rules with both conditions satisfied, we will also partially use the rules for whom only

one conditions is satisfied. In particular, if we have $x_1 > 0$ and $x_2 = 0$, then we will partially use the rules that correspond to the following 4 situations:

- if $x_1 > 0$ and $x_2 < 0$, then apply no control (because the pendulum is already moving in the direction towards the desired vertical position).
- if $x_1 > 0$ and $x_2 > 0$, then push the cart strongly.
- if $x_1 = 0$ and $x_2 = 0$, then apply no control (the pendulum is vertical and stays vertical).
- if $x_1 < 0$ and $x_2 = 0$, then push the cart into the opposite direction.

As a result, in addition to a rule that gives the desired control, we have 4 more rules, of which 1 controls in the right direction, 1 controls in the wrong direction, and 2 prescribe no control at all. If we combine the results of these 5 rules, we kind of "water down" the original control. Therefore, in this initial situation, if we want the deviation x_1 from the vertical axis to go to 0 as fast as possible, we better stay with the min and not use aggregation operations.

- On the second stage, at every moment of time (with a few exceptions), one of the following two reasoning methods turned out to be the best:
  - $f_\vee(a,b) = \max(a,b)$ and $f_\&(a,b) = \max(a+b-1,0)$;
  - $f_\vee(a,b) = \max(a,b)$ and $f_\&(a,b) = ab$.

  On this stage also, switching between these two reasoning methods occurred at seemingly random moments of time, and we could not trace any regularity in these switches.

In our first experiments, we only had an initial perturbation. As a result, the ideal control quickly decreased the perturbation to the level where it was practically 0. After that, we got what looked like an additional third stage, on which the deviations were so small that the computer precision interfered with our ability to compute the quality characteristics precisely and therefore, with our ability to make meaningful choices between different reasoning methods. As a result, on this third stage, we got a seemingly random sequence of choice of reasoning methods as supposedly "best". In real-life control, this never happens, because some perturbations are always there. So, in the consequent experiments, we added additional perturbations for $t > 0$. After we did that, the third stage disappeared, and the two-stage description turned out to be working OK.

6.8 Conclusions

The ideal control must consist of two stages:

- On the *first stage*, the main objective is to make the difference $x = X - X_0$ between the actual state X of the plant and its ideal state X_0 go to 0 as fast as possible.

Mathematically, it means that we have to minimize relaxation time. To achieve this goal, on this stage, we must use one if the following two reasoning methods:

- $f_\vee(a,b) = \min(a+b,1)$ and $f_\&(a,b) = \min(a,b)$;
- $f_\vee(a,b) = \min(a+b,1)$ and $\tilde{f}_\&(a,b) = p\min(a,b) + (1-p)(a+b)/2$.

The choice between these two methods depends on the particular circumstances: for one and the same plant, at different moments of time, different reasoning methods can lead to a better control.

- After we have already achieved the objective of the first stage, and the difference is close to 0, then the *second stage* starts. On this second stage, the main objective is to keep this difference close to 0 at all times. We do not want this difference to oscillate wildly, we want the dependency $x(t)$ to be as smooth as possible. So, mathematically, the objective is to minimize a functional $J(x(t))$ that describes the non-smoothness of the resulting trajectory. The smallest possible value of this function is achieved if we use one of the following two reasoning methods:
  - $f_\vee(a,b) = \max(a,b)$ and $f_\&(a,b) = \max(a+b-1,0)$, or
  - $f_\vee(a,b) = \max(a,b)$ and $f_\&(a,b) = ab$.

  On this stage also, for one and the same plant, at different moments of time, different reasoning methods can lead to a better (smoother) control.

6.9 Open problems

- Our results have been proved under two simplifying assumptions:
  - the plant is simple,
  - centroid defuzzification is used, and
  - for aggregation "and", no more than two conditions are allowed in each rule.

  It is therefore desirable to extend these results to the general case, i.e., to the case when:
  - plants are more realistic,
  - other defuzzification methods are allowed,
  - rules with three or more conditions are allowed.

  An empirical comparison of different switching techniques has been undertaken in [21–23]. This analysis seems to confirm our results in a more general framework than they are stated so, a theoretical explanation is in order.
- In this paper, we have found the optimal reasoning *methods* between which to switch. However, we did not produce any theoretical description of the *optimal switching time.* This is still an open problem.

Acknowledgments. This work was partially supported by NSF Grants No. CDA-

9015006 and No. EEC-9322370, NASA Grants No. 9-482 and NCC 2-275, MICRO State Programs No. 90-184 and 92-180, EPRI Agreement RP 8010-34, the BISC (Berkeley Initiative in Soft Computing) Program, and by a Grant No. PF90–018 from the General Services Administration (GSA), administered by the Materials Research Institute and the Institute for Manufacturing and Materials Management. The authors are greatly thankful to Lotfi Zadeh for encouraging this research, and to David Berleant, Ron Fearing, Bob Lea, Arnold Neumaier, Hung T. Nguyen, and Hideuki Takagi for valuable discussions.

6.10 References

[1] Berenji, H.R. "Fuzzy logic controllers". In: "An Introduction to Fuzzy Logic Applications in Intelligent Systems" (R. R. Yager, L. A. Zadeh. eds.), Kluwer Academic Publ., 1991.

[2] Blin, J.M. "Fuzzy relations in group decision theory", J. of Cybernetics, 4, 17–22, 1974.

[3] Blin, J.M., and A.B. Whinston. "Fuzzy sets and social choice". J. of Cybernetics, 3, 28–36, 1973.

[4] Buchanan, B.G., and E.H. Shortliffe. "Rule-based expert systems. The MYCIN experiments of the Stanford Heuristic Programming Project." Addison-Wesley, Reading, MA, Menlo Park, CA, 1984.

[5] Chang, S.S.L., and L.A. Zadeh. "On fuzzy mapping and control", IEEE Transactions on Systems, Man and Cybernetics, SMC-2, 30–34, 1972.

[6] D'Azzo, J.J., and C.H. Houpis. "Linear control system analysis and design: conventional and modern". Mc-Graw Hill, N.Y., 1988.

[7] Dubois, D., and H. Prade. "Fuzzy sets and systems: theory and applications." Academic Press, N.Y., London, 1980.

[8] Dubois, D., and H. Prade. "Possibility theory. An approach to computerized processing of uncertainty". Plenum Press, N.Y. and London, 1988.

[9] Hersch, H.M., and A. Caramazza. "A fuzzy-set approach to modifiers and vagueness in natural languages", J. Exp. Psychol.: General, 105, 254–276, 1976.

[10] Klir, G.J., and T.A. Folger. "Fuzzy sets, uncertainty and information". Prentice Hall, Englewood Cliffs, NJ, 1988.

[11] Kreinovich, V.; C. Quintana, R. Lea, O. Fuentes, A. Lokshin, S. Kumar, I. Boricheva, and L. Reznik. "What non-linearity to choose? Mathematical foundations of fuzzy control," Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control, Louisville, KY, 349–412, 1992.

[12] Kosko, B. “Neural networks and fuzzy systems”, Prentice Hall, Englewood Cliffs, NJ, 1992.

[13] Lea, R.N. “Automated space vehicle control for rendezvous proximity operations”. Telemechanics and Informatics, 5, 179–185, 1988.

[14] Lea, R.N.; Y.K. Jani, and H. Berenji. “Fuzzy logic controller with reinforcement learning for proximity operations and docking”. Proceedings of the 5th IEEE International Symposium on Intelligent Control, 2, 903–906, 1990.

[15] Lea, R.N., M. Togai, J. Teichrow, and Y. Jani. “Fuzzy logic approach to combined translational and rotational control of a spacecraft in proximity of the Space Station”. Proceedings of the 3rd International Fuzzy Systems Association Congress, 23–29, 1989.

[16] Lee, C.C. “Fuzzy logic in control systems: fuzzy logic controller.” IEEE Transactions on Systems, Man and Cybernetics, 20(2), 404–435, 1990.

[17] Mamdani, E.H. “Application of fuzzy algorithms for control of simple dynamic plant”, Proceedings of the IEE, 121(12), 1585–1588, 1974.

[18] Oden, G.C. “Integration of fuzzy logical information”, Journal of Experimental Psychology: Human Perception Perform., 3(4), 565–575, 1977.

[19] Savage, L.J. “The foundations of statistics”. Wiley, N.Y., 1954.

[20] Shortliffe, E.H. “Computer-based medical consultation: MYCIN”. Elsevier, N. Y., 1976.

[21] Smith, M.H. “Evaluation of performance and robustness of a parallel dynamic switching fuzzy system”, Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems, College Station, 118–121, 1992.

[22] Smith, M.H. “Parallel dynamic switching of reasoning methods in a fuzzy system”, Proceedings of the 2nd IEEE-FUZZ International Conference, San Francisco, CA, 1, 968–973, 1993.

[23] Smith, M.H., and H. Takagi, “Optimization of fuzzy systems by switching reasoning methods dynamically”, Fifth International Fuzzy Systems Association World Congress, Seoul, Korea, 1354–1357, 1993.

[24] Sugeno, M. (editor). “Industrial applications of fuzzy control”, North Holland, Amsterdam, 1985.

[25] Zadeh, L. “Fuzzy sets”, Information and control, 8, 338–353, 1965.

[26] Zimmerman, H.J. “Results of empirical studies in fuzzy set theory”. In: G.J. Klir (ed.), “Applied General System Research”, Plenum, N. Y., 303–312, 1978.

[27] Zimmermann, H.J. “Fuzzy set theory and its applications”, Kluwer, N.Y., 1991.