

Approximation of the Continuous Nilpotent Operator Class

József Dombi, Zsolt Gera

University of Szeged, Institute of Informatics
E-mail: {dombi | gera}@inf.u-szeged.hu

Abstract: In this paper we propose an approximation of the class of continuous nilpotent operators. The proof is based on one hand the approximation of the cut function, and on the other hand the representation theorem of operators with zero divisors. The approximation is based on sigmoid functions which are found to be useful in machine intelligence and other areas, too. The continuous nilpotent class of operators play an important role in fuzzy logic due to their good theoretical properties. Besides them this operator family does not have a continuous gradient. The main motivation was to have a simple and continuously differentiable approximation which ensures good properties for the operator.

Keywords: nilpotent operators, sigmoid function, approximation

1 Introduction

The nilpotent operator class (see e.g. [1], [2], [3]) is commonly used for various purposes. In the following we will consider only the continuous nilpotent operators. In this well known operator family the cut function (denoted by $[\square]$) plays a central role. We can get the cut function from x by taking the maximum of 0 and x and then taking the minimum of the result and 1. One can relax the restrictions of 0 and 1 to get the concept of the generalized cut function.

Definition 1 *Let the cut function be*

$$[x] = \min(1, \max(0, x)) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } 1 < x \end{cases} \quad (1)$$

Let the generalized cut function be

$$[x]_{a,b} = [(x-a)/(b-a)] = \begin{cases} 0 & \text{if } x < a \\ (x-a)/(b-a) & \text{if } a \leq x \leq b \\ 1 & \text{if } b < x \end{cases} \quad (2)$$

where $a, b \in \mathbb{R}$ and $a < b$.

Remark. We will use $[\cdot]$ for parentheses, too, e.g. $f[x]$ means $f([x])$.

As it can be seen from the representation theorem of the nilpotent class, which we will show later, all nilpotent operators are constructed using the cut function. The formulas of the Lukasiewicz conjunction, disjunction, implication and negation are the following:

$$\begin{aligned} c(x, y) &= [x + y - 1] \\ d(x, y) &= [x + y] \\ i(x, y) &= [y - x + 1] \\ n(x) &= 1 - x \end{aligned} \quad (3)$$

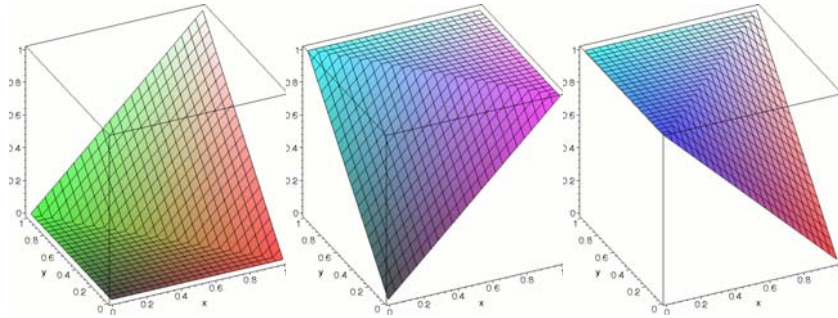


Figure 1

The truth tables of the Lukasiewicz conjunction, disjunction and implication

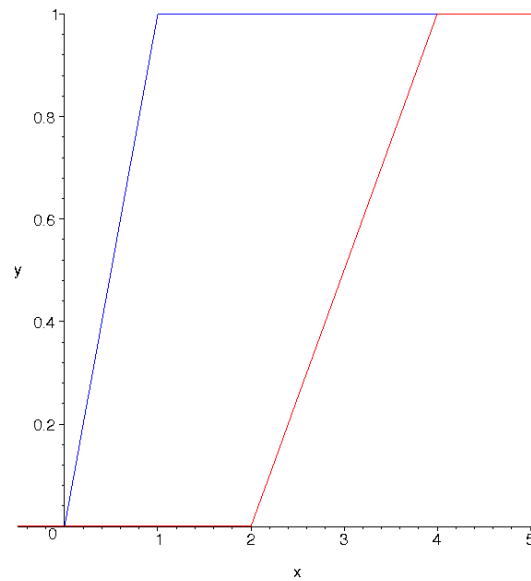


Figure 2

Two generalized cut functions

The truth tables of the former three can be seen on figure 1. The Lukasiewicz operator family used above has good theoretical properties. These are: the law of non-contradiction (that is the conjunction of a variable and its negation is always zero) and the law of excluded middle (that is the disjunction of a variable and its negation is always one) both hold, and the residual and material implications coincide. These properties make these operators to be widely used in fuzzy logic and to be the closest one to classic Boolean logic. Besides these good theoretical properties this operator family does not have a continuous gradient. So for example gradient based optimization techniques are impossible with Lukasiewicz operators. The root of this problem is the shape of the cut function itself.

2 Approximation of the Cut Function

A solution to above mentioned problem is a continuously differentiable approximation of the cut function, which can be seen on figure 3. In this section we'll construct such an approximating function by means of sigmoid functions. The reason for choosing the sigmoid function was that this function has a very important role in many areas. It is frequently used in artificial neural networks, optimization methods, economical and biological models.

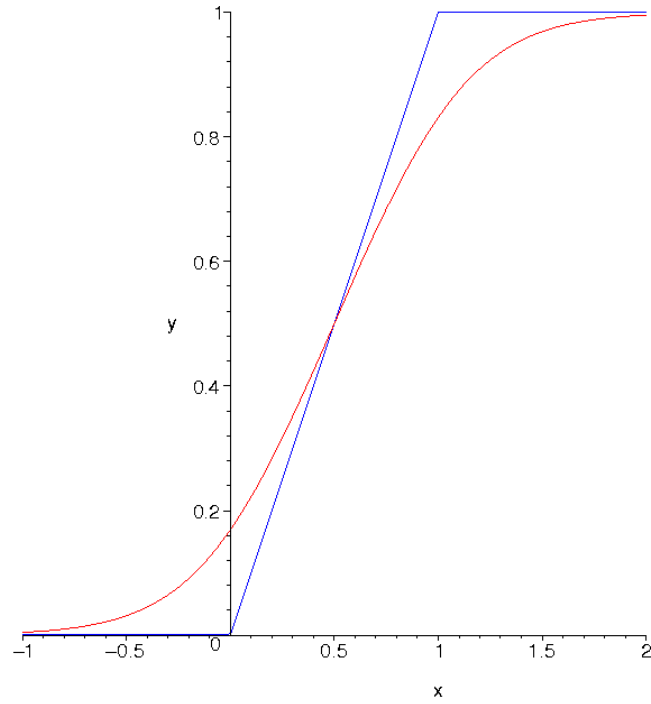


Figure 3
The cut function and its approximation

2.1 The Sigmoid Function

The sigmoid function (see figure 4) is defined as

$$\sigma_d^{(\beta)}(x) = \frac{1}{1 + e^{-\beta(x-d)}} \quad (4)$$

where the lower index d is omitted if 0.

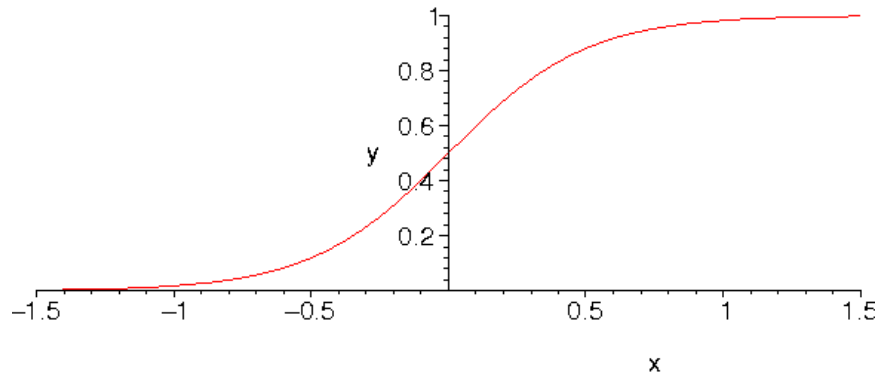


Figure 4
The sigmoid function with parameters $d=0$ and $\beta=4$

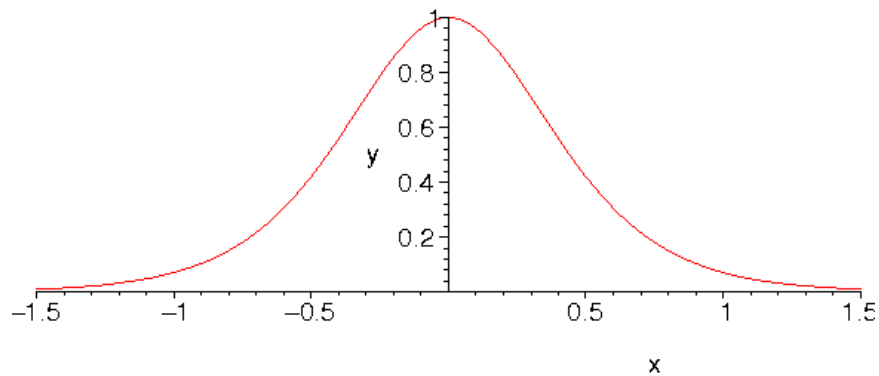


Figure 5
The first derivative of the sigmoid function

Let us examine some of its properties which will be useful later:

- its derivative can be expressed by itself (see figure 5):

$$\frac{\partial \sigma_d^{(\beta)}(x)}{\partial x} = \beta \sigma_d^{(\beta)}(x) \sigma_d^{(-\beta)}(x), \quad (5)$$

- its integral has the following form:

$$\int \sigma_d^{(\beta)}(x) dx = -\frac{1}{\beta} \ln \sigma_d^{(-\beta)}(x). \quad (6)$$

Because the sigmoid function is asymptotically 1 as x tends to infinity, the integral of the sigmoid function is asymptotically x (see figure 6).

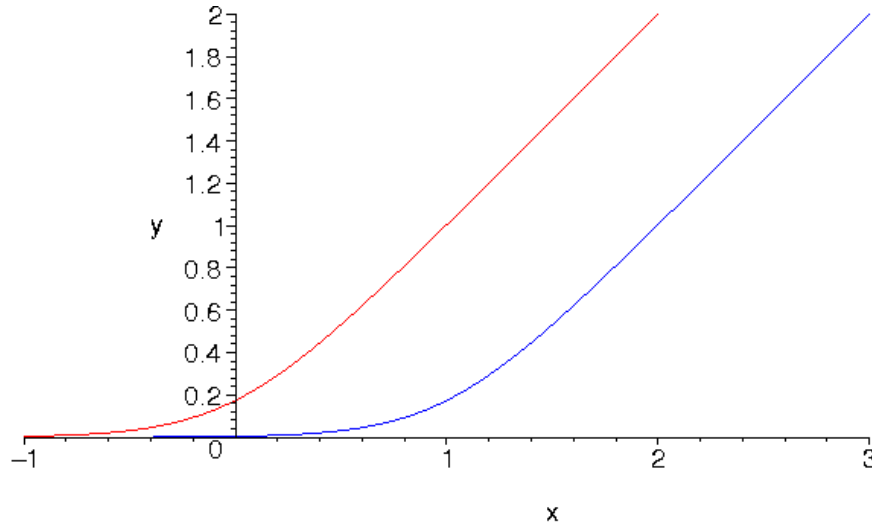


Figure 6
The integral of the sigmoid function, one is shifted by 1

2.2 The Squashing Function on the Interval [a,b]

In order to get an approximation of the generalized cut function, let us integrate the difference of two sigmoid functions, which are translated by a and b ($a < b$), respectively.

$$\begin{aligned} \frac{1}{b-a} \int \sigma_a^{(\beta)}(x) - \sigma_b^{(\beta)}(x) dx &= \frac{1}{b-a} \left(\int \sigma_a^{(\beta)}(x) dx - \int \sigma_b^{(\beta)}(x) dx \right) = \\ &= \frac{1}{b-a} \left(-\frac{1}{\beta} \ln \sigma_a^{(-\beta)}(x) + \frac{1}{\beta} \ln \sigma_b^{(-\beta)}(x) \right) \end{aligned} \quad (7)$$

After simplification we get the squashing function on the interval $[a,b]$:

Definition 2 Let the *interval squashing function* on $[a,b]$ be

$$S_{a,b}^{(\beta)}(x) = \frac{1}{b-a} \ln \left(\frac{\sigma_b^{(-\beta)}(x)}{\sigma_a^{(-\beta)}(x)} \right)^{1/\beta} = \frac{1}{b-a} \ln \left(\frac{1 + e^{\beta(x-a)}}{1 + e^{\beta(x-b)}} \right)^{1/\beta}. \quad (8)$$

The parameters a and b affect the placement of the interval squashing function, while the β parameter drives the precision of the approximation. We need to prove that $S_{a,b}^{(\beta)}(x)$ is really an approximation of the generalized cut function.

Theorem 3 Let $a, b \in \mathbb{R}$, $a < b$ and $\beta \in \mathbb{R}^+$. Then

$$\lim_{\beta \rightarrow \infty} S_{a,b}^{(\beta)}(x) = [x]_{a,b} \quad (9)$$

and $S_{a,b}^{(\beta)}(x)$ is continuous in x , a , b and β .

Proof. It is easy to see the continuity because $S_{a,b}^{(\beta)}(x)$ is a simple composition of continuous functions and because the sigmoid function has a range of $[0,1]$ the quotient is always positive.

In proving the limit we separate three cases, depending on the relation between a, b and x .

- Case 1 ($x < a < b$): Since $\beta(x-a) < 0$, so $e^{\beta(x-a)} \rightarrow 0$ and similarly $e^{\beta(x-b)} \rightarrow 0$. Hence the quotient converges to 1 if $\beta \rightarrow \infty$, and the logarithm of one is zero.
- Case 2 ($a \leq x \leq b$):

$$\begin{aligned} & \frac{1}{b-a} \ln \left(\lim_{\beta \rightarrow \infty} \left(\frac{1 + e^{\beta(x-a)}}{1 + e^{\beta(x-b)}} \right)^{1/\beta} \right) = \\ & = \frac{1}{b-a} \ln \left(\lim_{\beta \rightarrow \infty} \left(\frac{e^{\beta(x-a)} (e^{-\beta(x-a)} + 1)}{1 + e^{\beta(x-b)}} \right)^{1/\beta} \right) = \\ & = \frac{1}{b-a} \ln \left(\lim_{\beta \rightarrow \infty} \frac{e^{x-a} (e^{-\beta(x-a)} + 1)^{1/\beta}}{(1 + e^{\beta(x-b)})^{1/\beta}} \right) = \\ & = \frac{1}{b-a} \ln \left(e^{x-a} \lim_{\beta \rightarrow \infty} \frac{(e^{-\beta(x-a)} + 1)^{1/\beta}}{(1 + e^{\beta(x-b)})^{1/\beta}} \right). \end{aligned} \quad (10)$$

We transform the nominator so that we can take the e^{x-a} out of the limes. In the nominator $e^{-\beta(x-a)}$ remained which converges to 0 as well as $e^{\beta(x-b)}$ in the denominator so the quotient converges to 1 if $\beta \rightarrow \infty$. So as the result, the limit of the interval squashing function is $(x-a)/(b-a)$, which by definition equals to the generalized cut function in this case.

- Case 3 ($a < b < x$):

$$\begin{aligned} & \frac{1}{b-a} \ln \left(\lim_{\beta \rightarrow \infty} \left(\frac{1 + e^{\beta(x-a)}}{1 + e^{\beta(x-b)}} \right)^{1/\beta} \right) = \\ & = \frac{1}{b-a} \ln \left(\lim_{\beta \rightarrow \infty} \left(\frac{e^{\beta(x-a)} (e^{-\beta(x-a)} + 1)}{e^{\beta(x-b)} (e^{-\beta(x-b)} + 1)} \right)^{1/\beta} \right) = \\ & = \frac{1}{b-a} \ln \left(\lim_{\beta \rightarrow \infty} \frac{e^{x-a} (e^{-\beta(x-a)} + 1)^{1/\beta}}{e^{x-b} (e^{-\beta(x-b)} + 1)^{1/\beta}} \right) = \\ & = \frac{1}{b-a} \ln \left(\frac{e^{x-a}}{e^{x-b}} \lim_{\beta \rightarrow \infty} \frac{(e^{-\beta(x-a)} + 1)^{1/\beta}}{(e^{-\beta(x-b)} + 1)^{1/\beta}} \right). \end{aligned} \tag{11}$$

We do the same transformations as in the previous case but we take e^{x-b} from the denominator, too. After these transformations the remaining quotient converges to 1, so

$$\begin{aligned} \lim_{\beta \rightarrow \infty} S_{a,b}^{(\beta)}(x) &= \frac{1}{b-a} \ln \left(\frac{e^{x-a}}{e^{x-b}} \right) = \frac{1}{b-a} \ln(e^{x-a-(x-b)}) = \\ &= \frac{1}{b-a} \ln e^{b-a} = \frac{b-a}{b-a} = 1. \end{aligned} \tag{12}$$

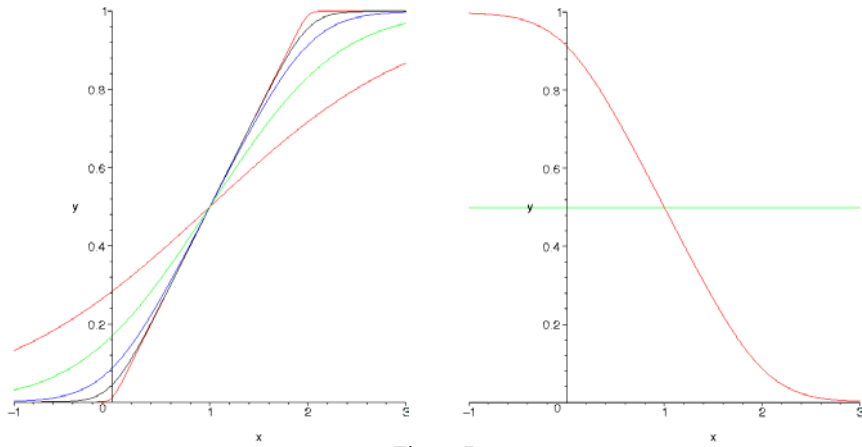


Figure 7

On the left image: the interval squashing function with an increasing β parameter ($a=0$ and $b=2$). On the right image: the interval squashing function with a zero and a negative β parameter

On figure 7 the interval squashing function can be seen with various β parameters. The following proposition states some properties of the interval squashing function.

Proposition 4

$$\lim_{\beta \rightarrow 0} S_{a,b}^{(\beta)}(x) = 1/2, \quad (13)$$

$$S_{a,b}^{(-\beta)}(x) = 1 - S_{a,b}^{(\beta)}(x). \quad (14)$$

As an another example, the Lukasiewicz conjunction is approximated with the interval squashing function on figure 8.

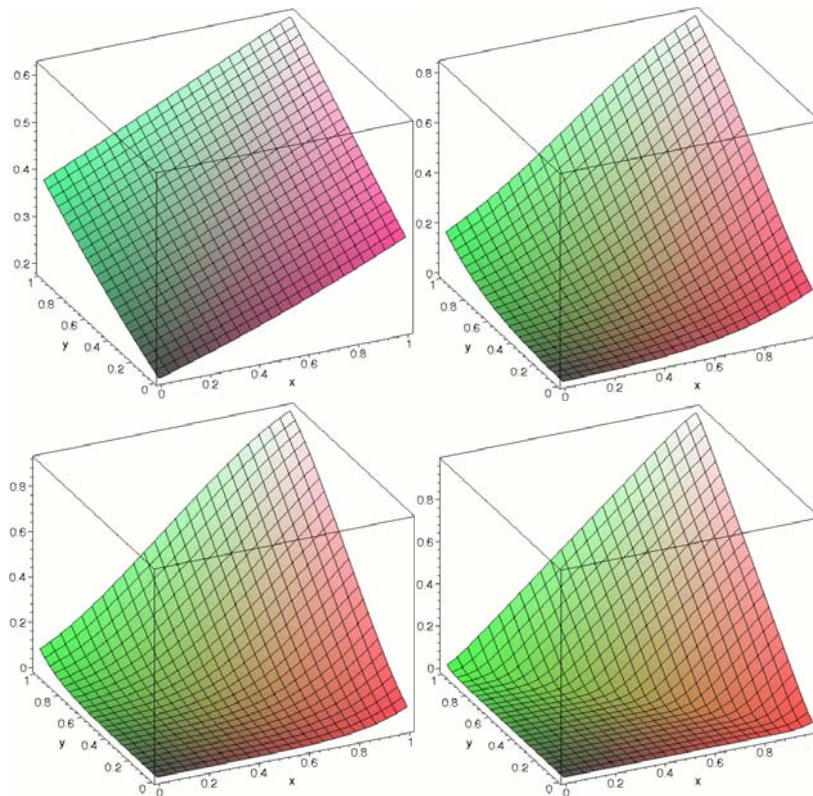


Figure 8

The approximation of the Lukasiewicz conjunction $[x+y-1]$ with β values 1,2,8 and 32

For further use, let us introduce an another form of the interval squashing function's formula. Instead of using parameters a and b which were the "bounds" on the x axis, from now on we'll use a and δ , where a gives the center of the squashing function and where δ gives its steepness. Together with the new

formula we introduce its pliant notation.

Definition 5 Let the *squashing function* be

$$\langle a <_{\delta} x \rangle_{\beta} = S_{a,\delta}^{(\beta)}(x) = \frac{1}{2\delta} \ln \left(\frac{\sigma_{a+\delta}^{(-\beta)}(x)}{\sigma_{a-\delta}^{(-\beta)}(x)} \right)^{1/\beta}, \quad (15)$$

where $a \in \mathbb{R}$ and $\delta \in \mathbb{R}^+$.

If the a and δ parameters are both 1/2 we will use the following pliant notation for simplicity:

$$\langle x \rangle_{\beta} = S_{1/2,1/2}^{(\beta)}(x), \quad (16)$$

which is the approximation of the cut function.

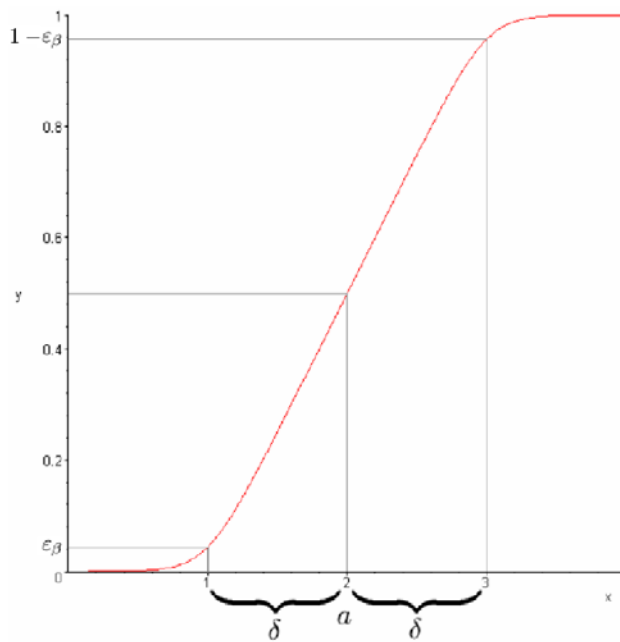


Figure 9

The meaning of $\langle a <_{\delta} x \rangle_{\beta}$

The inequality relation in the pliant notation refers to the fact that the squashing function can be interpreted as the truthness of the relation $a < x$ with decision level 1/2, according to a fuzziness parameter δ and an approximation parameter β (see figure 9).

The derivatives of the squashing function can be expressed by itself and sigmoid

functions:

$$\frac{\partial S_{a,\delta}^{(\beta)}(x)}{\partial x} = \frac{1}{2\delta} \left(\sigma_{a-\delta}^{(\beta)}(x) - \sigma_{a+\delta}^{(\beta)}(x) \right), \quad (17)$$

$$\frac{\partial S_{a,\delta}^{(\beta)}(x)}{\partial a} = \frac{1}{2\delta} \left(\sigma_{a+\delta}^{(\beta)}(x) - \sigma_{a-\delta}^{(\beta)}(x) \right), \quad (18)$$

$$\frac{\partial S_{a,\delta}^{(\beta)}(x)}{\partial \delta} = \frac{1}{2\delta} \left(\sigma_{a+\delta}^{(\beta)}(x) + \sigma_{a-\delta}^{(\beta)}(x) \right) + \frac{1}{\delta} S_{a,\delta}^{(\beta)}(x). \quad (19)$$

2.3 The Error of the Approximation

The squashing function approximates the cut function with an error. This error can be defined in many ways. We have chosen the following definition.

Definition 6 Let the approximation error of the squashing function be

$$\varepsilon_\beta = \langle 0 <_\delta (-\delta) \rangle = \frac{1}{2\delta} \ln \left(\frac{\sigma_\delta^{(-\beta)}(-\delta)}{\sigma_{-\delta}^{(-\beta)}(-\delta)} \right)^{1/\beta} \quad (20)$$

where $\beta > 0$.

Because of the symmetry of the squashing function $\varepsilon_\beta = 1 - \langle 0 <_\delta \delta \rangle$, see figure 9.

The purpose of measuring the approximation error is the following inverse problem: we want to get the corresponding β parameter for a desired ε_β error. We state the following lemma on the relationship between ε_β and β .

Lemma 7 Let us fix the value of δ . The following holds for e_β .

$$\varepsilon_\beta < c \cdot \frac{1}{\beta}, \quad (21)$$

where $c = \frac{\ln 2}{2\delta}$ is constant.

Proof.
$$\varepsilon_\beta = \frac{1}{2\delta\beta} \ln \left(\frac{1 + e^{\beta(-\delta+\delta)}}{1 + e^{\beta(-\delta-\delta)}} \right) = \frac{1}{2\delta\beta} \ln \left(\frac{2}{1 + e^{-2\delta\beta}} \right) =$$

$$= \frac{\ln 2}{2\delta\beta} - \frac{\ln(1 + e^{-2\delta\beta})}{2\delta\beta} < c \cdot \frac{1}{\beta} \quad (22)$$

So the error of the approximation can be upper bounded by $c \cdot \frac{1}{\beta}$, which means that by increasing parameter β , the error decreases by the same order of magnitude. ■

2.4 The Approximation of the Nilpotent Operator Class

The following theorems state that any continuous t-norm having zero divisors can be represented by the Lukasiewicz t-norm. In other words any nilpotent t-norm can be constructed using the Lukasiewicz t-norm and an appropriate automorphism of the unit interval. By these theorems the approximation of the cut function and the Lukasiewicz t-norm can be extended to the approximation of the whole class of continuous nilpotent operators.

We give the theorems using a different notation as stated in [4], especially for the cut function. By using $[\cdot]$ in the expressions, both the t-norm and the t-conorm case can be notated in the same way. The following lemma is needed for proving the theorems.

Lemma 8 *If c is a continuous t-norm such that $c(x, n(x)) = 0$ holds for all $x \in [0, 1]$ with a strict negation n then c is Archimedean.*

Proof. Suppose that c is not Archimedean. That is, there exists $x \in (0, 1)$ such that $c(x, x) = x$. If $x \leq n(x)$ then $x = c(x, x) \leq c(x, n(x)) = 0$, a contradiction since $x \in (0, 1)$. If $x > n(x)$ then, since c is a continuous function, there exists $y \leq x$ such that $n(x) = c(x, y)$. Then we have

$$n(x) = c(x, y) = c(c(x, x), y) = c(x, c(x, y)) = c(x, n(x)) = 0,$$

again a contradiction since $x \in (0, 1)$. Thus our proposition is proved. ■

Theorem 9 *A continuous t-norm c is such that $c(x, n(x)) = 0$ holds for all $x \in [0, 1]$ with a strict negation n if and only if there exists an automorphism f of the unit interval such that*

$$c(x, y) = f^{-1}[f(x) + f(y) - 1] \quad (23)$$

and

$$n(x) \leq f^{-1}(1 - f(x)). \quad (24)$$

Proof. (Necessity) According to the previous proposition, c is Archimedean. Thus

there exists a generator f_c of c such that $c(x, y) = f_c^{(-1)}(f_c(x) + f_c(y))$ (where $f_c^{(-1)}$ is the pseudoinverse of f_c) with $f_c(0) < \infty$. Define

$$f(x) = 1 - \frac{f_c(x)}{f_c(0)}. \quad (25)$$

Thus f is an automorphism of the unit interval. From (25) we have

$$f_c(x) = f_c(0) - f_c(0)f(x) \quad \text{and} \quad f_c^{(-1)}(x) = f^{-1}\left(1 - \frac{x}{f_c(0)}\right).$$

Using the above generator functional form of $c(x, y)$ we can go on as

$$\begin{aligned} c(x, y) &= f_c^{(-1)}(f_c(x) + f_c(y)) = f^{-1}[f_c(x) + f_c(y)] = \\ &= f^{-1}\left(1 - \frac{[f_c(0) - f_c(0)f(x) + f_c(0) - f_c(0)f(y)]}{f_c(0)}\right) = \\ &= f^{-1}[f(x) + f(y) - 1]. \end{aligned}$$

On the other hand, $c(x, n(x)) = 0$ is equivalent to $f(x) + f(n(x)) \leq 1$, whence we obtain the inequality $n(x) \leq f^{-1}(1 - f(x))$.

Proof of sufficiency is immediate. ■

We give the theorem for t -conorms without proof since it is very similar to the above mentioned.

Theorem 10 *A continuous t -conorm d satisfies condition $d(x, n(x)) = 1$ for all $x \in [0, 1]$ with a strict negation n if and only if there exists an automorphism g of the unit interval such that*

$$d(x, y) = g^{-1}[g(x) + g(y)] \quad (26)$$

and

$$n(x) \geq g^{-1}(1 - g(x)). \quad (27)$$

Using the above theorems we can state the following.

Theorem 11 *Every continuous nilpotent t -norm and t -conorm can be approximated by the squashing function in the following way:*

$$c(x, y) = f^{-1}\langle f(x) + f(y) - 1 \rangle_{\beta} \quad (28)$$

$$d(x, y) = g^{-1}\langle g(x) + g(y) \rangle_{\beta} \quad (29)$$

Proof. Because $\langle x \rangle_{\beta}$ approximates $[x]$ for all x if $\beta \rightarrow \infty$, the statement of the theorem is obvious from Theorem 9 and 10.

Conclusion

In this paper first we reviewed the cut function, which is the basis of the well known nilpotent operator class. This cut function is piecewise linear, hence it can not be continuously differentiated. We have constructed an approximation of the cut function (the squashing function) by means of sigmoid functions with good analytical properties, for example fast convergence and easy calculation. We have shown that all nilpotent operators can be approximated this way.

References

- [1] R. Ackermann. *An Introduction to Many-Valued Logics*. Dover, New York, 1967
- [2] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer, 1998
- [3] D. Mundici, R. Cignoli, I. M. L. D'Ottaviano. *Algebraic foundations of many-valued reasoning*. Trends in Logic, 7, 2000
- [4] J. Fodor, M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Theory and Decision Library. Kluwer, 1994
- [5] J. Dombi, Zs. Gera. *The Approximation of Piecewise Linear Membership Functions and Lukasiewicz Operators*. Fuzzy Sets and Systems (manuscript, submitted for publication)