

Gottschling, Andreas; Kreuter, Christof

**Working Paper**

## Approximation properties of the neuro-fuzzy minimum function

Research Notes, No. 99-3

**Provided in Cooperation with:**

Deutsche Bank Research, Frankfurt am Main

*Suggested Citation:* Gottschling, Andreas; Kreuter, Christof (1999) : Approximation properties of the neuro-fuzzy minimum function, Research Notes, No. 99-3, Deutsche Bank Research, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/40292>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Approximation Properties of the Neuro-Fuzzy Minimum Function**

Andreas Gottschling and Christof Kreuter

March 1999

## **Abstract**

The integration of fuzzy logic systems and neural networks in data driven nonlinear modeling applications has generally been limited to functions based upon the multiplicative fuzzy implication rule for theoretical and computational reasons. We derive a universal approximation result for the minimum fuzzy implication rule as well as a differentiable substitute function that allows fast optimization and function approximation with neuro-fuzzy networks.

**JEL:** C0; C2; C4; C6

**Keywords:** Fuzzy Logic, Neural Networks, Nonlinear Modeling, Optimization

**Andreas Gottschling, Christof Kreuter**

Deutsche Bank Research, Große Gallusstr. 10-14, 60311 Frankfurt, Germany

E-mail: andreas.gottschling@db.com, christof.kreuter@db.com

## Introduction

The integration of linguistic information in the form of fuzzy logic and statistical knowledge acquisition by neural networks has led to the emerging field of neuro-fuzzy systems. In the context of nonlinear modeling, this type of model combines some of the attractive features from each of the original concepts. In a neuro-fuzzy system expert knowledge can be used for the initialization of the parameterized nonlinear function, implemented as a feedforward network. Such a neuro-fuzzy network is based upon a particular type of nonlinear transformation, which is, as is the general case in neural networks, implemented at the hidden layer level. Specifically, the nonlinear structure has to satisfy the mathematical representation of a logical implication rule. The benefit of satisfying both, the fuzzy logic and the neural network conditions are:

1) enabling the use of various sophisticated data driven optimization techniques to improve on the potentially inaccurate or incomplete information provided by the expert.

2) gaining an insight into the information obtained from the data because the nonlinear model resulting from a statistical optimization of the neuro-fuzzy system retains *meaningful* parameters, contrary to many alternative nonlinear modeling approaches, which are often characterized as *black box* methods.

However, neuro-fuzzy modeling is severely limited by the narrow scope of admissible functional specifications. The vast majority of neuro-fuzzy applications use one and the same nonlinear transformation, namely the one associated with the multiplicative (product) implication structure. This is due to:

- 1) the lack of approximation-theoretic justification for alternative logical implication rules (IF-THEN rules)
- 2) computational convenience, since the differentiability of the networks is frequently lost when moving from the product rule to alternative implications.

The narrow scope of functional and interpretational variation, implied by the availability of only a single neuro-fuzzy specification, naturally limits its use. This is particularly unsatisfactory in economics and finance, given that interpretable nonlinear models constitute one of the few means to improve our understanding of the complex - and probably nonlinear - interaction mechanisms generating much of the observed empirical data.

To remedy these facts, we provide the theoretical basis for the empirical application of an alternative neuro-fuzzy system. In this system the nonlinear transformation corresponds to the minimum rule of implication. We first provide the necessary universal

approximation results<sup>1</sup> to allow consistent nonlinear function approximation with minimum-implication based neuro-fuzzy networks. Second, to overcome caveat 2, a differentiable extension of the minimum function is derived. This allows the application of fast optimization algorithms to the neuro-fuzzy network. Several simulations illustrate the intuition behind these results.

## Universal Approximation

### Definitions

*Feedforward Neural Network:* For any  $r \in \mathbb{N}$  let  $A$  be an affine transformation of  $x \in \mathbb{R}^r$ . Using  $\Psi: \mathbb{R}^r \rightarrow \mathbb{R}$  (called *combination* or *implication* function) and  $g: \mathbb{R} \rightarrow \mathbb{R}$  (called *transfer* or *activation* function), define  $f: \mathbb{R}^r \rightarrow \mathbb{R}$ :

$$f(x) = \sum_{j=1}^q \beta_j \cdot \Psi[g(A_{j1}(x)), \dots, g(A_{jr}(x))] \quad (1)$$

with  $\beta_j \in \mathbb{R}, q = 1, 2, \dots$

$f(x)$  is called a feedforward neural network. This definition allows for complex, multivariate nonlinear transformations at the hidden layer level while retaining the additively separable structure underlying key aspects in the neural network literature.

*Fuzzy set:* Let  $U = \mathbb{R}^r$ ; a set  $A \subset U$  is a fuzzy set if its set membership function is multivalued, e.g.  $\mu_A(x): U \rightarrow [0, 1]$ , where  $\mu_A(x)$  is the "membership grade of point  $x$  in  $A$ ". As a contrast, in the case of an ordinary or "crisp" set  $A$  the function  $\mu_A(x): U \rightarrow \{0, 1\}$ , i.e. it is only bivalued, meaning that either  $x$  belongs to  $A$  or it does not. Any crisp point  $x \in U$  can be "fuzzified". For example one possible fuzzification of the crisp point  $\sqrt{2}$  ( $\in \mathbb{R}$ ), could be achieved by any continuous probability density function  $f$ , centered at  $\sqrt{2}$  and normalized such that  $f(\sqrt{2}) = 1$ . This transformation *smears* out  $x$  over a whole range with varying membership grade.

*Fuzzy Rule:* A fuzzy IF-THEN rule is of the form: IF  $x_1$  is  $A_1$  and ... and  $x_n$  is  $A_n$  THEN  $y$  is  $B$ , where " $x_k$  is  $A_k$ " stands for the degree of membership of  $x_k$  in  $A_k$ ;  $A_k$  and  $B$  are fuzzy sets.

*Fuzzy Logic System*<sup>2</sup>: a mapping from  $\mathbb{R}^r \rightarrow \mathbb{R}$  described by one of the following functional forms:

<sup>1</sup> A functional family has universal approximation characteristic if arbitrarily exact approximation of any function in the universe of interest is possible.

<sup>2</sup> Limited to the fuzzy logic systems of interest in the context of this paper. Neither exhaustive nor all-encompassing. Many alternative fuzzy logic systems known to the authors are hereby excluded.

1. Product rule:

$$f(x) = \sum_{j=1}^q \beta_j \cdot \mu_{A_{j,1}}(x_1) \cdot \dots \cdot \mu_{A_{j,r}}(x_r) \quad (2)$$

2. Minimum rule:

$$f(x) = \sum_{j=1}^q \beta_j \cdot \min[\mu_{A_{j,1}}(x_1), \dots, \mu_{A_{j,r}}(x_r)] \quad (3)$$

The difference in logical implication between the two rules can be illustrated in the following example: The *probability* of a joint failure of a two independent component system is given by the product (rule 1) of the individual probabilities to fail. The *possibility* of system failure is given as soon as one of the components fails, thus the minimum (rule 2) of the two probabilities yields this information, since the stronger component does have to fail for the joint event to occur. This is equivalent to stating that a combination of events *can* occur exactly if the least likely event of all events occurs. Hence, taking the probability of the failure of the strongest link of any system as an estimate of the risk is obviously the most conservative approach for any risk calculation as it corresponds to the extreme case of perfect correlation.

As seen above, in a neuro-fuzzy network each logical implication corresponds to a particular functional form of the nonlinear transformation  $\Psi$ . In general, all logically interpretable functions are constrained by the structural requirements<sup>3</sup> for admissible  $\Psi$ ; the desirable feature of meaningful parameters hence acts as an important determinant of the function approximator.

To fit a nonlinear model such as a neuro-fuzzy network to empirical data, apart from the interpretability, one requires functional consistency. The neuro-fuzzy network has to be capable of adequately capturing arbitrary nonlinear functions<sup>4</sup>, which could be underlying the data generating process. This universal approximation property obviously depends on the properties of both, the implication function  $\Psi: \mathfrak{R}^r \rightarrow \mathfrak{R}$  and the nonlinear transfer function  $g: \mathfrak{R} \rightarrow \mathfrak{R}$ , because they jointly determine the nonlinearity at the hidden layer level. Consistency is given only for  $\Psi$  being the product implication with Gaussian transfer functions  $g$  (e.g. see [5]) or a power thereof [2].

Hence we need to determine sufficient conditions on the transfer function  $g$ , such that there exists a neuro-fuzzy systems with the minimum implication rule, which can approximate an arbitrary continuous function to any desired degree of accuracy.

### Theorem 1

Let  $\mathbb{C}(\mathfrak{R}^r)$  denote the space of continuous functions from  $\mathfrak{R}^r \rightarrow \mathfrak{R}$  and  $K \subseteq \mathbb{C}(\mathfrak{R}^r)$  a compact subspace. Define

$$f(x) \equiv \sum_{j=1}^q \beta_j \cdot \min_k [g(a_{o,k} + a_{1,k} \cdot x_k)] \quad (4)$$

with  $k \in \{1, \dots, r\}$   $q \in \{1, 2, \dots\}$  and  $g: \mathfrak{R} \rightarrow \mathfrak{R}$  integrable, bounded and continuous almost everywhere s.th.  $g(x) = g(-x)$ ,  $g(x) < g(y)$  for  $|x| > |y|$  with  $\int_{\mathfrak{R}^r} g(x) dx \neq 0$ . Then

for any  $F(x) \in \mathbb{C}(\mathfrak{R}^r)$  and for any  $\varepsilon > 0 \exists f(x)$  s.th.  
 $\sup_K \|F(x) - f(x)\| < \varepsilon$ .

### Proof

Let  $\|\cdot\|_p$  denote a p-norm<sup>5</sup>. Based upon [3], it has been established that the functional family<sup>6</sup> defined by:

$$f_p(x, q) = \sum_{j=1}^q \beta_j \cdot g\left(\frac{1}{\sigma} \cdot \|x - z_j\|_p\right) \quad (5)$$

with  $x, z_j \in \mathfrak{R}^r$ ,  $\beta_j \in \mathfrak{R}$ ,  $\sigma \in \mathfrak{R}^+$  is dense on  $K \subseteq \mathbb{C}(\mathfrak{R}^r)$  for any  $p \in [1, \infty)$ , if  $g$  is integrable, bounded and continuous almost everywhere and it holds that:  $\int_{\mathfrak{R}^r} g(x) dx \neq 0$ <sup>7</sup>. One can thus construct a dense

functional family on the compact subspace  $K \subseteq \mathbb{C}(\mathfrak{R}^r)$  for countably many  $p$ . Since

$$\lim_{p \rightarrow \infty} \sqrt[p]{|x_1|^p + |x_2|^p + \dots + |x_r|^p} = \max\{|x_1|, |x_2|, \dots, |x_r|\} \quad (6)$$

(as shown e.g. in [1]) it follows that (5) converges uniformly to

$$f_\infty(x, q) = \sum_{j=1}^q \beta_j \cdot g\left(\frac{1}{\sigma} \max\{|x_1 - z_{j,1}|, \dots, |x_r - z_{j,r}|\}\right) \quad (7)$$

when  $p \rightarrow \infty$ . In order to establish the universal approximation property on the compact subspace  $K \subseteq \mathbb{C}(\mathfrak{R}^r)$  for equation (7), we show that for any  $\varepsilon > 0$  and arbitrary  $F(x) \in \mathbb{C}(\mathfrak{R}^r) \exists f_\infty(x, q)$  s.th.

$$\sup_K |F(x) - f_\infty(x, q)| < \varepsilon. \quad (8)$$

The following conditions are fulfilled:

(I) for any  $F(x) \in \mathbb{C}(\mathfrak{R}^r)$  and for any  $\eta > 0 \exists q^* \in [1, \infty)$  s.th.  $\forall q' \geq q^*$  it holds:

$$\sup_K |F(x) - f_p(x, q')| < \eta$$

for any fixed  $p \in [1, \infty)$ . This follows from the consistency of  $f_p(x, q')$ , derived in [3].

<sup>5</sup>  $\|x\|_p \equiv \sqrt[p]{|x_1|^p + |x_2|^p + \dots + |x_r|^p}$

<sup>6</sup> This type of function is known as radial basis function and/or Kernel estimator in the literature.

<sup>7</sup> E.g. this holds among others for any continuous probability density function  $g$ .

<sup>3</sup> For more details see e.g. [5].

<sup>4</sup> We limit ourselves to continuous functions for the sake of exposition. Extension to  $L_2$  follows naturally [5].

(II) for preset values of  $p, p'$  and  $q' \in [1, \infty)$  and for any  $\varphi > 0 \exists q^* \in [1, \infty)$  s.th.  $\forall q \geq q^*$  it holds:

$$\sup_K |f_p(x, q') - f_{p'}(x, q)| < \varphi$$

follows as a special case from (I).

(III) for any fixed  $q \in [1, \infty)$  and for any  $\delta > 0$   
 $\exists p^* \in [1, \infty)$  s.th.  $\forall p' \geq p^*$  it holds that:

$$\sup_K |f_{p'}(x, q) - f_\infty(x, q)| < \delta$$

follows from equation (7).

The repeated application of the triangle inequality to the left-hand side of (8) yields:

$$\begin{aligned} \sup_K |F(x) - f_\infty(x, q)| &\leq \sup_K |F(x) - f_p(x, q')| + \\ &\sup_K |f_p(x, q') - f_{p'}(x, q)| + \sup_K |f_{p'}(x, q) - f_\infty(x, q)| \end{aligned} \quad (9)$$

where each of the right-hand side terms is arbitrarily small because they obey conditions I, II and III, respectively:

$$\sup_K |F(x) - f_\infty(x, q)| \leq \eta + \varphi + \delta \equiv \varepsilon \quad (10)$$

Thus one can always find a  $q$  such that the left-hand side is arbitrarily small. This result establishes the universal approximation of systems such as (7). To apply this result to minimum-implication rule fuzzy systems note that for any  $g: \mathfrak{R} \rightarrow \mathfrak{R}$  s.th.  $g(x) = g(-x)$  and  $g(x) < g(y)$  for  $|x| > |y|$  a functional equivalence between (4) and (7) follows from

$$\min_k \{h(x_k)\} = h\left(\max_k \{|x_k|\}\right) \quad (11)$$

by setting  $a_{0,k} = -z_{j,k}/\sigma$  and  $1/\sigma = a_{l,k} \forall k, k \in \{1, \dots, r\}$ . This establishes the density of minimum implication fuzzy logic systems via a functional equivalence relation to the radial basis function. Hence subject to the conditions on  $g(x)$  all admissible radial basis function kernels (5) can be used for consistent modeling with the minimum implication rule as well. Since the previous consistency results for fuzzy logic approximators were limited to the Gaussian density function [5] and powers thereof [2], this theorem provides a significant extension to the scope of consistent fuzzy modeling.

### Properties of the Fuzzy Minimum System

Given the theoretical justification for the use of the minimum implication in the nonlinear approximation context, it is interesting to investigate its properties. The first question obviously concerns the domain of application, i.e. in what type of problem is a minimum implication system more suitable than a product rule system?

The differences between product and minimum implication rule are best illustrated in the form of graphical representations (Figures 1-2), defined as the neuro-fuzzy system output displayed on a 2D surface over the input quantities  $x_1$  and  $x_2$  ( $q=1$ ).

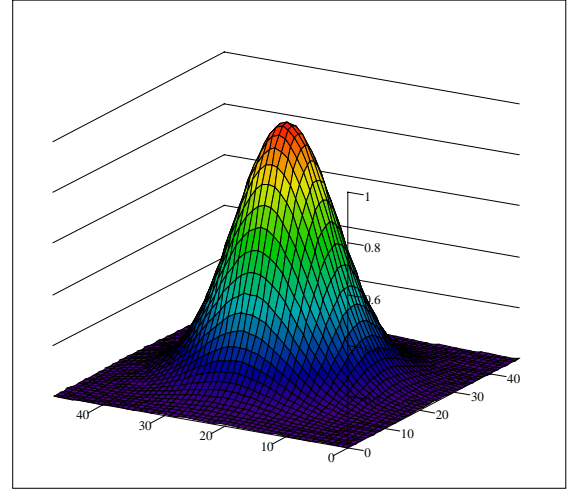


Figure 1: Graphical representation of the product implication rule with Gaussian transfer function.

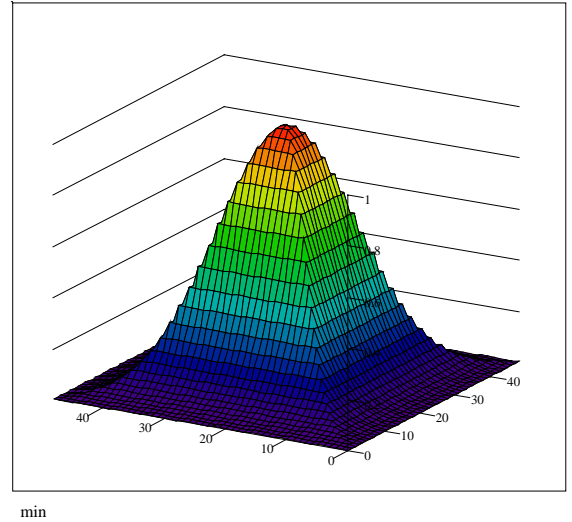


Figure 2: Graphical representation of the minimum implication rule with Gaussian transfer function.

The two graphs show significant differences in the structure of their level sets. It becomes apparent: the less convex the level sets of any target function are the more appropriate the minimum function becomes. Furthermore, consider the shape of admissible membership functions. As stated above Gaussians and their powers are so far the only choice in case of the product rule system, however any symmetric unimodal function, centered at zero and strictly monotonic on either side of its maximum constitutes an acceptable membership function for the minimum rule (Figure 3).

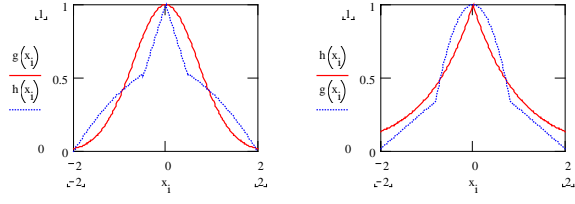


Figure 3: Acceptable membership functions for the minimum implication rule.

As an example for function approximation, we consider a dynamical systems problem where the present value of  $x$  depends on functions of the most recent changes and functions of the most recent levels such as<sup>8</sup>:

$$x_t = -1 + (x_{t-1} - x_{t-2})^{2/3} + (x_{t-1} + x_{t-2})^{2/3} \quad (12)$$

which displays chaotic behavior for start values  $x_1 = -0.3$  and  $x_2 = -0.1$ .

A comparison of the goodness of fit<sup>9</sup> for both, the minimum and the product implication rule yields:

	Product Rule	Minimum Rule
Mean Squared Error	8.1	3.1

Table 1: Goodness of Fit of Product and Minimum Rule for time series (12).

While the figures indicate that in this particular case the minimum function proves superior to the product, it is important to note that the minimum function also introduces a problem which becomes severe in complex applications: any cost function based upon it will not be differentiable because the minimum function is not differentiable itself; this prevents the usage of efficient (gradient) optimization methods. To illustrate this point we contrast convergence times of fitting 600 data points from series (12), once with the differentiable product implication rule and then with the minimum implication rule for the following optimization algorithms [4]:

1. the quasi-Newton algorithm by Broydon, Fletcher, Goldfarb, and Shanno (BFGS)
2. the conjugate gradient method as formulated by Fletcher, Reeves, Polak, and Ribiere
3. Powell's modified conjugate directions method
4. the simplex simulated annealing (SSA) algorithm.

<sup>8</sup> Similar problems arise in the control literature.

<sup>9</sup> 600 data points from time series (12) are fitted using one hidden node systems with each implication rule using the simplex simulated annealing (SSA) algorithm of [4].

Table 2 reports the obtained average of convergence times for the different combinations of implication rule and optimization procedure. Note that due to the lack of differentiability the gradient algorithms 1. and 2. can not be applied to the minimum implication system.

Algorithm	Product	Minimum
1.Quasi Newton	189	no gradient
2.Conj. Grad.	1888	no gradient
3.Powell	4993	7322
4.SSA	7642	15121

Table 2: Average convergence time in microseconds.

It is evident that for the product implication rule the gradient algorithms significantly outperform algorithms 3. and 4. Hence - even if the minimum implication constitutes a potentially superior<sup>10</sup> structure - the lack of differentiability is a major hindrance for any application of the minimum implication system on large data sets and/or high dimensional problems. For such applications a differentiable substitute function for the minimum implication is derived in the following paragraph. This provides (almost) the same functional properties as the minimum function, augmented with with an analytical gradient for efficient optimization.

### A Differentiable Quasi-Minimum Function

The differentiable substitute function is derived in two steps. First the bivariate case is considered and then we show that this argument can be recursively extended to any finite dimensional multivariate  $x$ .

Rewriting (4) for 2-dimensions (Figure 2):

$$f(x) = \sum_{j=1}^q \beta_j \cdot \min[g(a_{o,1} + a_{1,1} \cdot x_1), g(a_{o,2} + a_{1,2} \cdot x_2)] \quad (13)$$

Substituting  $y_i = g(a_{o,i} + a_{1,i} \cdot x_i)$ ,  $i \in \{1,2\}$ ,

the minimum function can be expressed as:

$$\min(y_1, y_2) = \frac{1}{2} \cdot (y_1 + y_2 - |y_1 - y_2|). \quad (14)$$

Since the absolute value function fails to be differentiable, any function based upon it inherits this property.

Consider the following substitution for the absolute value function  $|z|$ :

$$\Psi(z, \alpha) = \frac{1}{\alpha} \ln(\cosh(\alpha z) \cdot (\tanh^2(\alpha z) + 1)) \quad (15)$$

This substitution has the following properties:

<sup>10</sup> If the target surface features non-convex level sets such as those displayed by equation (12) or cusps, the minimum implication rule yields better results.

**Lemma 1**

$$1. \quad \lim_{\alpha \rightarrow \infty} \Psi(z, \alpha) = |z| \quad (16)$$

$$2. \quad \Psi(0, \alpha) = 0 \quad (17)$$

$$3. \quad \lim_{\alpha \rightarrow \infty} \frac{d\Psi(z, \alpha)}{dz} = \begin{cases} 1 & \text{for } z > 0 \\ -1 & \text{for } z < 0 \end{cases} \quad (18)$$

$$4. \quad \Psi(z, \alpha) \in \mathbb{C}(2) \quad (19)$$

$$5. \quad \Psi(z, \alpha) \leq |z| \quad \forall \alpha > 0 \quad (20)$$

**Proof**

1. For  $z = 0$  the result is evident from property 2. For  $z \neq 0$  we show that  $\lim_{\alpha \rightarrow \infty} \Psi(z, \alpha) - |z| = 0$  using L'Hopital's rule. The ratio of the derivatives w.r.t.  $\alpha$  yields:

$$\tanh(\alpha \cdot z) \cdot z + 2 \cdot \tanh(\alpha \cdot z) \cdot \frac{1 - \tanh^2(\alpha \cdot z)}{1 + \tanh^2(\alpha \cdot z)} \cdot z - |z|.$$

The central term vanishes because  $\lim_{v \rightarrow \pm\infty} \tanh^2(v) = 1$ . Since  $\lim_{v \rightarrow \infty} \tanh(v) = 1$  and  $\lim_{v \rightarrow -\infty} \tanh(v) = -1$  the first term cancels with  $|z|$ , hence the property follows.

2. by inspection

3. Given that the first derivative w.r.t.  $z$  of  $\Psi(z, \alpha)$  is

$$\frac{d\Psi(z, \alpha)}{dz} = \frac{\tanh(\alpha \cdot z)(3 - \tanh^2(\alpha \cdot z))}{1 + \tanh^2(\alpha \cdot z)},$$

and noting that  $\lim_{v \rightarrow \infty} \tanh(v) = 1$  and  $\lim_{v \rightarrow -\infty} \tanh(v) = -1$ ,

$$\text{we find: } \lim_{\alpha \rightarrow \infty} \frac{d\Psi(z, \alpha)}{dz} = \begin{cases} 1 & \text{for } z > 0 \\ -1 & \text{for } z < 0 \end{cases}.$$

$$4. \quad \frac{d^2\Psi(z, \alpha)}{dz^2} = \alpha \left( \frac{(1 - \eta^2) \cdot (3 - 6\eta^2 - \eta^4)}{(1 + \eta^2)^2} \right)$$

with  $\eta = \tanh(\alpha \cdot z)$ . The denominator is bounded away from zero for every term. Thus the second derivative as the sum and product of continuous functions is continuous itself.

5. To see that

$$\frac{1}{\alpha} \ln(\cosh(\alpha z) \cdot (\tanh^2(\alpha z) + 1)) - z \leq 0$$

for all  $z > 0$  and all  $\alpha > 0$ , some algebra yields:

$$-2 \cdot \alpha z \leq 0,$$

which is evident for  $\alpha > 0$  and  $z > 0$ . The same steps can be done for  $z \leq 0$  by changing the sign of the inequality. The proposed function serves as a parametric substitute for the absolute value function (Figure 4).

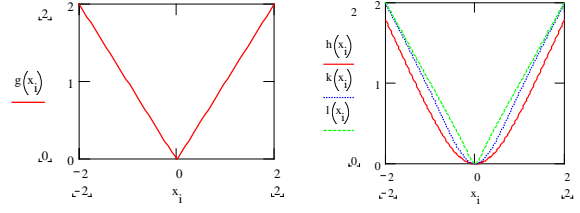


Figure 4: Absolute value function  $|z|$  (left) and its substitute  $\Psi(z, \alpha)$  for  $\alpha = 0.5, 1.0$  and  $5.0$  (right)..

For the case of  $q=1$ , the substitution yields:

$$f(x) = \frac{\beta_1}{2} \cdot [y_1 + y_2 - \frac{\ln(\cosh(\alpha(y_1 - y_2)) \cdot (\tanh^2(\alpha(y_1 - y_2)) + 1))}{\alpha}] \quad (21)$$

for the right-hand side of equation (13). This constitutes a 2-dimensional differentiable quasi-minimum function (Figure 5).

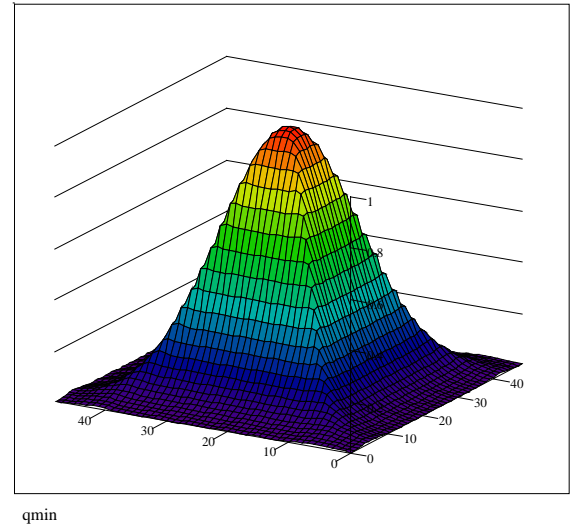


Figure 5: The quasi-minimum implication rule with Gaussian transfer function ( $\alpha=10$ )

To extend this argument to  $r$  dimensions, note that

$$\min\{y_1, \dots, y_r\} = \min\{y_i, \min\{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_r\}\} \quad (22)$$

for any  $i \in \{1, \dots, r\}$ , i.e. the minimum function can be recursively applied pairwise to any number and permutation of the arguments without changing the result [1]. Thus by recursive pairwise application of

$$\min(y_i, y_j) = \frac{1}{2} \cdot (y_i + y_j - |y_i - y_j|) \quad (23)$$

with

$$y_j = \min\{y_1, \dots, y_{i-1}\} \quad (24)$$

we can extend the minimum function to any finite

dimension  $r$ . Consequently, by substituting the differentiable quasi-absolute value function  $\Psi(y_i, y_j, \alpha)$  for  $|y_i - y_j|$  in each recursion leads to a differentiable quasi-minimum function defined from  $\mathcal{R}^r \rightarrow \mathcal{R}$ .

### Optimization Implications

A comparison of the average convergence times based upon time series (12) of the quasi-minimum function (using  $\alpha=1$ ) relative to the true minimum function is given in Table 3.

Algorithm	Minimum	Quasi-Min.
1.Quasi Newton	No gradient	309
2.Conj. Grad.	No gradient	4393
3.Powell	7322	11868
4.SSA	15121	35861

Table 3: Average convergence time in microseconds.

It turns out that, although the average convergence time for the quasi-minimum function is somewhat lower than for the product implication rule (Table 2), the applicability of the efficient gradient techniques (i.e. methods 1. and 2.) for the quasi-minimum function provides significant improvements in convergence speed with respect to the true minimum function. The relative performance differentials increase with the number of data points and the number of hidden nodes.

A remaining question pertains to the sensitivity of these results to  $\alpha$ . Figure 6 reports the mean squared error (MSE) of the same experiment for varying values of  $\alpha$ .

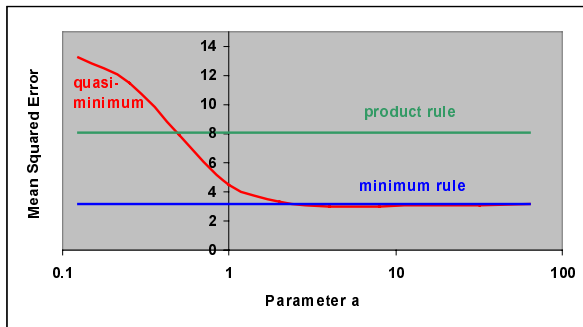


Figure 6:  $\alpha$ -dependence of the mean squared error for product, minimum, and quasi-minimum function.

For a large range of  $\alpha$  the approximation with the quasi-minimum yields the same results as that of the true minimum function. In this range the mean squared error of the minimum is significantly lower than that of the product approximation. While this is not intended to serve as an exhaustive evaluation of the properties of quasi-minimum approximation, these encouraging

results warrant further investigation and application of the minimum and the quasi-minimum neuro-fuzzy functions.

### Conclusions

This paper provides both, the theoretical basis and some practical hints for the application of neuro-fuzzy networks using the minimum-implication rule. A gap among the available universal approximation results for fuzzy logic systems is closed by theorem 1, which links the particular form of nonlinear transformations via a functional equivalence relation to existing consistent function approximators in the literature. Since this can be achieved by only one symmetry and one monotonicity constraint, the resulting admissible class of nonlinear *transfer* (or *activation* or *membership grade*) functions is significantly larger than the corresponding class for the product implication rule. Furthermore, we show how the inherently not differentiable minimum function can be replaced by an asymptotically equivalent implication rule, which enables significantly faster data driven optimization.

### References

- [1] Barner, M., Flohr, F. 1983 *Analysis II*. Publisher: De Gruyter, Berlin
- [2] Gottschling, A. 1997 *Structural Relationships between Feedforward Neural Networks, Fuzzy Logic Systems and Radial Basis Functions*. First Chapter of *Three Essays in Neural Networks and Financial Prediction*. PhD Thesis, UC San Diego.
- [3] Park, J. and Sandberg, I.W. 1991. *Universal Approximation using Radial-Basis-Function Networks*. Neural Computation 3, 246-257.
- [4] Press, W. et al. 1995. *Numerical Recipes in C*. Cambridge University Press.
- [5] Wang, L.X. 1992. *Fuzzy Systems are Universal Approximators*. Proceedings of the IEEE International Conference on Fuzzy Systems. 1163 - 1170.



## **Recent Publications:**

A.Gottschling, C. Kreuter, P.Cornelius

“The Reaction of Exchange Rates and Interest Rates to News Releases”,

RN-98-2, June 1998

Dong Heon Kim

“Another Look at Yield Spreads: Monetary Policy and the Term Structure of Interest Rates“,

RN-98-3, October 1998

Jeffrey Sachs

“Creditor Panics: Causes and Remedies“,

RN-98-4, November 1998

Namwon Hyung

“Linking Series Generated at Different Frequencies and its Applications“

RN-99-1, December 1998

Yungsook Lee

“The Federal Funds Market and the Overnight Eurodollar Market“

RN-99-2, January 1999

**© 1999. Publisher: Deutsche Bank AG, DB Research, D-60272 Frankfurt am Main, Federal Republic of Germany, editor and publisher, all rights reserved. When quoting please cite „Deutsche Bank Research“.**

The information contained in this publication is derived from carefully selected public sources we believe are reasonable. We do not guarantee its accuracy or completeness, and nothing in this report shall be construed to be a representation of such a guarantee. Any opinions expressed reflect the current judgement of the author, and do not necessarily reflect the opinion of Deutsche Bank AG or any of its subsidiaries and affiliates. The opinions presented are subject to change without notice. Neither Deutsche Bank AG nor its subsidiaries/affiliates accept any responsibility for liabilities arising from use of this document or its contents. Deutsche Bank Securities Inc. has accepted responsibility for the distribution of this report in the United States under applicable requirements. Deutsche Bank AG London and Morgan Grenfell & Co., Limited, both being regulated by the Securities and Futures Authority, have respectively, as designated, accepted responsibility for the distribution of this report in the United Kingdom under applicable requirements.