Implicative Fuzzy Associative Memories

Peter Sussner and Marcos Eduardo Valle

***Abstract*—Associative neural memories are models of biological phenomena that allow for the storage of pattern associations and the retrieval of the desired output pattern upon presentation of a possibly noisy or incomplete version of an input pattern. In this paper, we introduce implicative fuzzy associative memories (IFAMs), a class of associative neural memories based on fuzzy set theory. An IFAM consists of a network of completely interconnected Pedrycz logic neurons with threshold whose connection weights are determined by the minimum of implications of presynaptic and postsynaptic activations. We present a series of results for autoassociative models including one pass convergence, unlimited storage capacity and tolerance with respect to eroded patterns. Finally, we present some results on fixed points and discuss the relationship between implicative fuzzy associative memories and morphological associative memories.**

***Index Terms*—Associative memories, convergence, fuzzy Hebbian learning, fuzzy neural networks, fuzzy relations, fuzzy systems, morphological associative memories, storage capacity, tolerance with respect to noise.**

I. Introduction

GENERALLY speaking, a memory is a system with three functions or stages: 1) Recording: storing the information; 2) Preservation: keeping the information safely; 3) Recall: retrieving the information [1]. Research in psychology has shown that the human brain recalls by association, that is, the brain associates the recalled item with a piece of information or with another item [2]. The focus in this paper is on neural *associative memory* (AM) systems. Neural AMs are not only capable of storing and preserving associative information but can also be used to perfectly retrieve an item from incomplete or noisy information. These features make neural AMs a research area with applications in different fields of science, such as image processing, pattern recognition, and optimization [3]–[7].

Research on neural associative memories originated in the 1950s with the investigations of Taylor and Steinbuch on matrix associative memories [8], [9]. In 1972, Anderson, Kohonen, and Nakano introduced, independently, the *linear associative memory* (LAM) where correlation or hebbian learning is used to synthesize the synaptic weight matrix [10]–[12]. Perfect recall can be guaranteed by imposing an orthogonality condition on the stored patterns. The *optimal linear associative memory* (OLAM), which employs the projection recording recipe, is not subject to this constraint [13]–[15]. Although the OLAM exhibits a better storage capacity than the LAM model, this type of associative memory also has low noise tolerance due to the fact that it represents a linear model.

Manuscript received September 27, 2004; revised May 2, 2005. This work was supported in part by CNPq under grant numbers 142196/2003-7 and 303362/2003-0.

The authors are with the Institute of Mathematics, Statistics, and Scientific Computing, the State University of Campinas, São Paolo, Brazil (e-mail: sussner@ime.unicamp.br).

Digital Object Identifier 10.1109/TFUZZ.2006.879968

In 1982, Hopfield introduced a dynamic autoassociative memory model for bipolar patterns [16]. The *Hopfield net* revived research efforts in the area of neural networks due to the following attractive features: ease of implementation in hardware; characterization in terms of an energy function; variety of applications [17]. On the downside, the Hopfield net suffers from a low absolute storage capacity of approximately $0.15n$ patterns, where n is the length of the patterns. Several other models that appeared since then can be viewed as extensions of the Hopfield model [4]. For example, the *bidirectional associative memory* (BAM) generalizes the Hopfield model to include the heteroassociative case [18], [19]. The BAM retains the low absolute storage capacity of the Hopfield net. A significant improvement in storage capacity is achieved by the *exponencial capacity associative memory* (ECAM) [20]. Note, however, that the ECAM stores the information in a nondistributive manner in view of the fact that the columns of the synaptic weight matrix consist of the original patterns.

The associative memory models we mentioned above represent traditional semilinear neural networks. In each layer, a matrix-vector product of linear algebra is computed followed by the application of a possibly nonlinear activation function. In contrast to semilinear AM models, morphological and fuzzy associative memory systems employ a nonlinear matrix vector product at each layer and usually no activation function is needed [21]–[23]. A biological motivation for the nonlinear operations that are performed in morphological associative memories was given by Ritter and Urcid [24].

The earliest attempt to use fuzzy set theory to describe an associative memory was Kosko's *fuzzy associative memory* (FAM) [23]. This single-layer feedforward neural net is described in terms of a nonlinear matrix vector product called max-min or max-product composition, and the synaptic weight matrix is given by fuzzy Hebbian learning. Kosko's FAM model exhibits a large amount of crosstalk that leads to a very low storage capacity, i.e., one rule per FAM matrix. Chung and Lee generalized the FAM model using $\max$-t compositions and pointed out that a perfect recall of multiple rules per FAM matrix is possible if the input fuzzy sets are normal and $\max$-t orthogonal [25]. The *implicative fuzzy associative memory* (IFAM) model introduced in this paper does not encounter any of these difficulties. The great difference between the IFAM and Chung-Lee's *generalized fuzzy associative memory* (GFAM) model lies in the learning rule that was introduced in [26], [27]. In view of the implication used in this rule, we will speak of *implicative fuzzy learning*. Junbo *et al.* applied this learing rule to the particular case of max–min composition yielding the max-min fuzzy associative memory [27]. Liu adapted the max-min fuzzy associative memory by introducing a threshold into the model [28].

In this paper, we generalize implicative fuzzy learning to include any $\max$-t composition based on a continuous t-norm.

1063-6706/$20.00 © 2006 IEEE

We analyze the IFAM model and provide comparisons with others models by means of examples drawn from the literature [28], [30]. In the autoassociative case, we speak of *autoassociative fuzzy implicative memory* (AFIM). We show that the AFIM exhibits unlimited storage capacity, one-pass convergence, and tolerance with respect to erosive noise. Other associative memories models with unlimited storage capacity include the *morphological associative memory* (MAM) [21], [22], [29]. This paper shows that morphological associative memories with threshold can be viewed as a special case of IFAMs if the fundamental memory set is fuzzy and if the synaptic weight matrix is contained in $[0,1]^{m\times n}$.

The organization of the paper is as follows. Section II presents the neuro-fuzzy models, the network topology and the learning rules that will be used throughout the paper. Section III introduces the IFAM model and the dual IFAM model. This section also contains a comparison of the IFAM with other FAM models by means of an example from the literature and an application as a fuzzy rule-based system for a forecasting problem [28], [30]. Section IV describes characteristics such as convergence, storage capacity, noise tolerance and fixed points of AFIMs. We also show that every fixed point of a binary AFIM is given by a lattice polynomial in the original patterns. Section V reveals the relationship between the IFAM and the morphological associative memory model. This paper finishes with concluding remarks in Section VI. Some of the proofs of the theorems are provided in Appendix.

II. Background Information and Learning Rules for Neuro-Fuzzy Systems

Implicative fuzzy associative memories are fuzzy neural networks endowed with Pedrycz's logic based neurons with threshold [31], [32]. If $x_1, x_2, \ldots, x_n$ are the input signals then the output is given by the following equation:

$$y_i = \bigvee_{j=1}^{n} (w_{ij} \mathrm{~t~} x_j) \vee \theta_i, \qquad \text{for } i = 1, 2, \ldots, m. \tag{1}$$

A dual model is given in terms of the equation

$$y_i = \bigwedge_{j=1}^{n} (m_{ij} \mathrm{~s~} x_j) \wedge \vartheta_i, \qquad \text{for } i = 1, 2, \ldots, m. \tag{2}$$

Here, the symbols w_{ij} and m_{ij} denote the synaptic weights and θ_i and ϑ_i denote the thresholds (or biases) for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Here, $\vee$ and $\wedge$ represent the maximum and the minimum operations, and the symbol t represents a continuous t-norm and s represents a continuous s-norm. Note that the thresholds θ_i and ϑ_i can be interpreted as synaptic weights connected to a constant input x_0. That is, $\theta_i = w_{i0} \mathrm{~t~} x_0$, where $x_0 = 1$ in the first equation, and $\vartheta_i = m_{i0} \mathrm{~s~} x_0$, where $x_0 = 0$ in the second equation. We will use this fact to calculate the threshold value.

These two equations describe single layer feedforward fuzzy neural networks, the only architecture considered in this paper. In view of the nonlinearity of triangular norms and co-norms, we refrain from using any activation functions.

A network with m neurons described by (1) and (2) can be expressed using products of matrices [32]. For a matrix $A \in [0,1]^{m\times p}$ and $B \in [0,1]^{p\times n}$, the matrix $C = A \circ B \in [0,1]^{m\times n}$, called the *max-t composition of* A and B, is defined by the equation

$$c_{ij} = \bigvee_{k=1}^{p} (a_{ik} \mathrm{~t~} b_{kj}). \tag{3}$$

Similarly, the matrix $D = A \bullet B \in [0,1]^{m\times n}$ called the *min-s product composition of* A and B, is defined by the equation

$$d_{ij} = \bigwedge_{k=1}^{p} (a_{ik} \mathrm{~s~} b_{kj}). \tag{4}$$

Let $*$ denote a commutative, binary operation such as the maximum and the minimum. Given a scalar α and matrices A and B of the same size, the matrices $C = \alpha * A$ and $D = A * B$ are defined in terms of the equations $c_{ij} = \alpha * a_{ij}$ and $d_{ij} = a_{ij} * b_{ij}$, respectively.

In matrix form, (1) and (2) become

$$\mathbf{y} = (W \circ \mathbf{x}) \vee \boldsymbol{\theta} \tag{5}$$

and

$$\mathbf{y} = (M \bullet \mathbf{x}) \wedge \boldsymbol{\vartheta} \tag{6}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in [0,1]^n$ is the input pattern, $W = (w_{ij}) \in [0,1]^{m\times n}$ and $M = (m_{ij}) \in [0,1]^{m\times n}$ are the synaptic weight matrices, $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T \in [0,1]^m$ and $\boldsymbol{\vartheta} = [\vartheta_1, \vartheta_2, \ldots, \vartheta_m]^T \in [0,1]^m$ are the thresholds, and $\mathbf{y} = [y_1, y_2, \ldots, y_m]^T \in [0,1]^m$ is the output. If we interpret the threshold as a synaptic weight connected with a constant input, then (5) and (6) become

$$\mathbf{y} = [\boldsymbol{\theta} \quad W] \circ \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} = W_\theta \circ \mathbf{x}' \tag{7}$$

and

$$\mathbf{y} = [\boldsymbol{\vartheta} \quad M] \bullet \begin{bmatrix} 0 \\ \mathbf{x} \end{bmatrix} = M_\vartheta \bullet \mathbf{x}''. \tag{8}$$

Here, $W_\theta = [\boldsymbol{\theta}\ W]$ and $M_\vartheta = [\boldsymbol{\vartheta}\ M]$ are matrices in $[0,1]^{m\times(n+1)}$ and $\mathbf{x}' = [1\ \mathbf{x}]^T$ and $\mathbf{x}'' = [0\ \mathbf{x}]^T$ are vectors with $n+1$ components.

$\mathrm{Max}-t$ and $\min$-s compositions have interesting properties such as associativity and monotonicity [33], [34]. Moreover, the following distributive laws are valid for all matrices A, B and C of appropriate sizes:

$$A \circ (B \vee C) = (A \circ B) \vee (A \circ C) \tag{9}$$

$$A \bullet (B \wedge C) = (A \bullet B) \wedge (A \bullet C). \tag{10}$$

There is an elegant duality between $\max$-t and $\min$-s compositions. Let $\bar{R} = (\bar{r}_{ij}) \in [0,1]^{m\times n}$ be the complement matrix of $R = (r_{ij}) \in [0,1]^{m\times n}$ defined by

$$\bar{r}_{ij} = 1 - r_{ij} \tag{11}$$

for all i, j. If $A \in [0,1]^{m\times k}$ and $B \in [0,1]^{k\times n}$ are fuzzy matrices, then

$$\overline{A \circ B} = \bar{A} \bullet \bar{B}. \tag{12}$$

Equation (12) reveals that the $\max\text{-}t$ composition and the $\min\text{-}s$ composition are dual operations if the t-norm and s-norm used in these compositions are dual operations as well [34], [35]. Recall that if t is a t-norm, then we obtain

$$x\,s\,y = 1-(1-x)\mathrm{t}(1-y) = \overline{(\bar{x}\;t\;\bar{y})} \tag{13}$$

as the dual s-norm for every $x, y \in [0,1]$.

Example 2.1: The *minimum* $(\wedge)$ is a t-norm and its dual s-norm is the *maximum* $(\vee)$. The dual s-norm of the *product* is the *probabilistic sum* given by

$$x\,\mathrm{s}\,y = x+y-xy \tag{14}$$

for $x, y \in [0,1]$. The *t-norm of Lukasiewicz* (or bounded product) and its dual s-norm, called *s-norm of Lukasiewicz* (or bounded sum), are defined as follows:

$$x\,\mathrm{t}\,y = 0 \vee (x+y-1) \tag{15}$$
$$x\,\mathrm{s}\,y = 1 \wedge (x+y). \tag{16}$$

A. Implicative Fuzzy Learning

Hebb's postulate of learning states that the synaptic weight change depends on the input as well as the output activation [36]. In implicative fuzzy learning, the weights are adjusted according to a certain fuzzy implication. Thus, if x_j is the input and y_i is the output then the synaptic weight changes according to the equation

$$\Delta w_{ij} = (x_j \Rightarrow y_i). \tag{17}$$

Recall that a *fuzzy implication* is a mapping $\Rightarrow: [0,1]^2 \longrightarrow [0,1]$ that generalizes the crisp implication and satisfies the properties of monotonicity in the second argument, dominance of falsity, and neutrality of truth [34]. In this paper, we focus on a special type of fuzzy implication, namely the so-called R-implication associated with a continuous t-norm [35]. Recall that an *R-implication* is a mapping $\Rightarrow_R: [0,1]^2 \longrightarrow [0,1]$ given by

$$(x \Rightarrow_R y) = \sup\{z \in [0,1] : x\,\mathrm{t}\,z \leq y\}. \tag{18}$$

Example 2.2: Equations (19)–(21) represent R-implications associated with the minimum, the product, and the Lukasiewicz t-norm given by (15), respectively

$$\text{Gödel: } x \Rightarrow_R y = \begin{cases} 1, & x \leq y \\ y, & x > y \end{cases} \tag{19}$$
$$\text{Goguen: } x \Rightarrow_R y = 1 \wedge \left(\frac{y}{x}\right) \tag{20}$$
$$\text{Lukasiewicz: } x \Rightarrow_R y = 1 \wedge (1-x+y). \tag{21}$$

Suppose that we want to store the fundamental memory set $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1,\ldots,p\}$, where $\mathbf{x}^\xi \in [0,1]^n$ and $\mathbf{y}^\xi \in [0,1]^m$, using a synaptic weight matrix $W = (w_{ij}) \in [0,1]^{m\times n}$. Let $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p] \in [0,1]^{n\times p}$ denote the matrix whose columns are the input patterns and let $Y = [\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^p] \in [0,1]^{m\times p}$ denote the matrix whose columns are the output patterns. Since the minimum operation corresponds to the "and" conjunction, the matrix $W = (w_{ij}) \in [0,1]^{m\times n}$ is given by the minimum of the synaptic weight matrices for each pair $(\mathbf{x}^\xi, \mathbf{y}^\xi)$, i.e.,

$$w_{ij} = \bigwedge_{\xi=1}^{p} \left(x_j^\xi \Rightarrow y_i^\xi\right). \tag{22}$$

In other words, the synaptic weight matrix W is synthetized according the following rule:

$$W = Y \circledast X^T. \tag{23}$$

Here, the product $C = A \circledast B \in [0,1]^{m\times n}$ of the matrices $A \in [0,1]^{m\times k}$ and $B \in [0,1]^{k\times n}$ is computed as follows:

$$c_{ij} = \bigwedge_{l=1}^{k} (b_{lj} \Rightarrow a_{il}). \tag{24}$$

We refer to the rule given by (23) as *implicative fuzzy learning*. In the special case where a R-implication is used in (24), we speak of *R-implicative fuzzy learning* and we denote the matrix product using the symbol $\circledast_R$.

The following theorem holds true for the synaptic weight matrix W given by R-implicative fuzzy learning [37]:

Theorem 2.1: The synaptic weight matrix $W = Y \circledast_R X^T$ given by R-implicative fuzzy learning is the greatest solution of the fuzzy matrix inequality

$$W \circ X \leq Y. \tag{25}$$

In particular, if the fuzzy matrix equation $W \circ X = Y$ has a solution then $W = Y \circledast_R X^T$ is also a solution and this solution is the greatest one.

Implicative fuzzy learning will be used to train a fuzzy neural network described by (5). In view of the fact that $\circ$ and $\bullet$ are dual operations, we also define *dual implicative fuzzy learning* that will be used to train a network described by (6).

B. Dual Implicative Fuzzy Learning

We define a *dual fuzzy implication* as a mapping $\underline{\Rightarrow} : [0,1]^2 \longrightarrow [0,1]$ that satisfies the following relationship:

$$x \underline{\Rightarrow} y = \overline{(\bar{x} \Rightarrow \bar{y})} \qquad \forall\, x, y \in [0,1]. \tag{26}$$

Despite its name, a dual implication does generally not constitute an implication as defined in [34] and [35]. Applying (13) and (18) to (26) yields

$$x \underline{\Rightarrow}_R y = \inf\{z \in [0,1] : x\,\mathrm{s}\,z \geq y\} \tag{27}$$

for every $x, y \in [0,1]$. According to (27), a dual R-implication corresponds to a certain continuous s-norm.

Example 2.3: Equations (28)–(30) represent the dual R-implication associated with maximum, probabilistic sum, and Lukasiewicz s-norm, respectively

$$\text{Dual Gödel: } x \overline{\Rightarrow}_R y = \begin{cases} 0, & x \geq y \\ y, & x < y \end{cases} \tag{28}$$

$$\text{Dual Goguen: } x \overline{\Rightarrow}_R y = 0 \vee \left(\frac{y - x}{1 - x} \right) \tag{29}$$

$$\text{Dual Lukasiewicz: } x \overline{\Rightarrow}_R y = 0 \vee (y - x). \tag{30}$$

In *dual implicative fuzzy learning*, the matrix of synaptic weights M is synthesized as follows:

$$m_{ij} = \bigvee_{\xi=1}^{p} \left(x_j^{\xi} \overline{\Rightarrow} y_i^{\xi} \right). \tag{31}$$

The matrix M can be expressed as the $\bar{\circledast}$-product of the matrices $Y \in [0,1]^{m \times p}$ and $X^T \in [0,1]^{p \times n}$. In general, if $A \in [0,1]^{m \times k}$ and $B \in [0,1]^{k \times n}$ then the entries of $C = A \bar{\circledast} B \in [0,1]^{m \times n}$ can be determined as follows:

$$c_{ij} = \bigvee_{l=1}^{k} (b_{lj} \overline{\Rightarrow} a_{il}). \tag{32}$$

We will use the symbol $\bar{\circledast}_R$ for the special case where a dual R-implication is used in (32). The synaptic weight matrix $M = Y \bar{\circledast}_R X^T$ given by dual R-implicative fuzzy learning is the least solution of the fuzzy matrix inequality $M \bullet X \geq Y$ [37]. Finally, using the definitions of the products "$\circledast$" and "$\bar{\circledast}$" and the duality relationship of (26), we infer that

$$\overline{(Y \circledast X^T)} = \bar{Y} \bar{\circledast} \bar{X}^T. \tag{33}$$

III. Introduction to Implicative Fuzzy Associative Memories

An associative memory paradigm may be formulated as an input–output system that is able to store different patterns pairs $(\mathbf{x}, \mathbf{y})$ [4]. Mathematically speaking, the associative memory design problem can be stated as follows: Given a finite set of fundamental memories $\{(\mathbf{x}^{\xi}, \mathbf{y}^{\xi}) : \xi = 1, \ldots, p\}$, determine a mapping G such that $G(\mathbf{x}^{\xi}) = \mathbf{y}^{\xi}$ for all $\xi = 1, \ldots, p$. Furthermore, the mapping G should be endowed with a certain tolerance with respect to noise, i.e., $G(\tilde{\mathbf{x}}^{\xi})$ should equal $\mathbf{y}^{\xi}$ for noisy versions $\tilde{\mathbf{x}}^{\xi}$ of $\mathbf{x}^{\xi}$.

In neuro-fuzzy associative memories, the mapping G is described by a fuzzy neural network and the process used to find G becomes a learning rule. If R-implicative fuzzy learning is used to synthesize the synaptic weight matrix, we obtain an IFAM. Similarly, if dual R-implicative fuzzy learning is used, we obtain a dual IFAM. Note that there exist infinitely many IFAMs and dual IFAMs induced by triangular norms and co-norms. A particular IFAM is specified by its t-norm or R-implication and a particular dual IFAM is specified by its s-norm or dual R-implication.

The input and output patterns of an IFAM or a dual IFAM are fuzzy sets. Thus, the inputs $\mathbf{x}^{\xi}$ and the outputs $\mathbf{y}^{\xi}$, $\xi = 1, \ldots, p$, are interpreted as fuzzy sets in a finite universe of discourse $[0,1]^n$ and $[0,1]^m$.

We will only discuss neuro-fuzzy networks described by (5) or (6) or, equivalently, by (7) or (8). Determining G results in an estimation problem [34], because we need to determine the relation (matrix W_θ or M_ϑ) that satisfies (7) and (8) for each pair $(\mathbf{x}^{\xi}, \mathbf{y}^{\xi})$. It is known from Theorem 2.1 and its dual formulation that $W_\theta = Y \circledast_R (X')^T$ and $M_\vartheta = Y \bar{\circledast}_R (X'')^T$, where

$$(X')^T = \begin{bmatrix} \mathbf{1}_{1 \times p} \\ X \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \mathbf{x}^1 & \mathbf{x}^2 & \ldots & \mathbf{x}^p \end{bmatrix} \tag{34}$$

and

$$(X'')^T = \begin{bmatrix} \mathbf{0}_{1 \times p} \\ X \end{bmatrix} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ \mathbf{x}^1 & \mathbf{x}^2 & \ldots & \mathbf{x}^p \end{bmatrix} \tag{35}$$

are solutions of $W_\theta \circ X' \leq Y$ and $M_\vartheta \bullet X'' \geq Y$ [37]. Computing the threshold $\boldsymbol{\theta}$ as the first column of W_θ we obtain

$$\theta_i = \bigwedge_{\xi=1}^{p} \left(1 \Rightarrow_R y_i^{\xi} \right) = \bigwedge_{\xi=1}^{p} y_i^{\xi}. \tag{36}$$

Computing W from W_θ, we obtain $W = Y \circledast_R X^T$. Similarly, the threshold $\boldsymbol{\vartheta}$ is given by the equation

$$\vartheta_i = \bigvee_{\xi=1}^{p} \left(0 \overline{\Rightarrow}_R y_i^{\xi} \right) = \bigvee_{\xi=1}^{p} y_i^{\xi} \tag{37}$$

and $M = Y \bar{\circledast}_R X^T$. Note that the threshold $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ do not depend on the t-norm or s-norm used in the neuro-fuzzy model. Finally, as a consequence of Theorem 2.1, we have the following.

Corollary 3.1: If there exist W_0 and $\boldsymbol{\theta}_0$ such that $\mathbf{y}^{\xi} = (W_0 \circ \mathbf{x}^{\xi}) \vee \boldsymbol{\theta}_0$ for all $\xi = 1, 2, \ldots, p$ then $W_0 \leq W = Y \circledast_R X^T$, $\boldsymbol{\theta}_0 \leq \boldsymbol{\theta} = \bigwedge_{xi=1}^{p} \mathbf{y}^{\xi}$ and $\mathbf{y}^{\xi} = (W \circ \mathbf{x}^{\xi}) \vee \boldsymbol{\theta}$ holds true for all $\xi = 1, \ldots, p$.

Using the duality relationships given by (11), (12), and (33), we can also prove that if there exist M_0 and $\boldsymbol{\vartheta}_0$ such that $\mathbf{y}^{\xi} = (M \bullet \mathbf{x}^{\xi}) \wedge \boldsymbol{\vartheta}$ then $M_0 \geq M = Y \bar{\circledast}_R X^T$, $\boldsymbol{\vartheta}_0 \geq \boldsymbol{\vartheta} = \bigvee_{\xi=1}^{p} \mathbf{y}^{\xi}$ and $\mathbf{y}^{\xi} = (M \bullet \mathbf{x}^{\xi}) \wedge \boldsymbol{\vartheta}$ holds true for all $\xi = 1, \ldots, p$.

Summarizing, IFAMs are neuro-fuzzy networks described by

$$\mathbf{y} = (W \circ \mathbf{x}) \vee \boldsymbol{\theta} \tag{38}$$

where the synaptic weight matrix and the threshold are given by

$$W = Y \circledast_R X^T \text{ and } \boldsymbol{\theta} = \bigwedge_{\xi=1}^{p} \mathbf{y}^{\xi}. \tag{39}$$

Here, the implication used in (24) to compute the product $Y \circledast_R X^T$ is the R-implication associated with the t-norm of max-t composition of (38). Similarly, dual IFAMs are described by

$$\mathbf{y} = (M \bullet \mathbf{x}) \wedge \boldsymbol{\vartheta} \tag{40}$$

where

$$M = Y \bar{\circledast}_R X^T \text{ and } \boldsymbol{\vartheta} = \bigvee_{\xi=1}^{p} \mathbf{y}^{\xi}. \tag{41}$$

TABLE I
SET OF INPUT AND OUTPUT PAIRS USED IN EXAMPLE 3.1

ξ	$\mathbf{x}^\xi$	$\mathbf{y}^\xi$
1	$[0.5, 0.5, 0.4, 0.4, 0.3]^T$	$[0.5, 0.6, 0.3]^T$
2	$[0.1, 0.3, 0.3, 0.4, 0.4]^T$	$[0.5, 0.6, 0.4]^T$
3	$[0.8, 0.4, 0.6, 0.7, 0.4]^T$	$[0.6, 0.8, 0.4]^T$
4	$[0.3, 0.4, 0.4, 0.3, 0.4]^T$	$[0.5, 0.6, 0.4]^T$
5	$[0.6, 0.4, 0.7, 0.7, 0.5]^T$	$[0.7, 0.7, 0.5]^T$
6	$[0.1, 0.1, 0.2, 0.2, 0.1]^T$	$[0.5, 0.6, 0.3]^T$
7	$[0.7, 0.2, 0.4, 0.3, 0.2]^T$	$[0.5, 0.7, 0.3]^T$
8	$[0.8, 0.4, 0.3, 0.4, 0.2]^T$	$[0.5, 0.8, 0.3]^T$

Note that the dual R-implication that is used to form the product $Y \bar{\circledast}_R X^T$ is defined in terms of the same co-norm as the one employed in (40).

From now on, we will only discuss the IFAM model. Note that every statement concerning the IFAM model yields a corresponding dual statement concerning the dual IFAM model in view of (11)–(13) and (33). Specifically, we obtain the corresponding dual statement from the statement about the IFAM model by replacing minimum with maximum, t-norm with s-norm, implication with dual implication, the product $\circ$ with $\bullet$, $\circledast_R$ with $\bar{\circledast}_R$, and *vice versa*.

Example 3.1: An example provided by Liu indicates that the absolute storage capacity of the max–min FAM with threshold exceeds the storage capacity of the FAM of Junbo *et al.*, Kosko's FAM, and the generalized FAM of Chung-Lee [23], [25], [27], [28].

Consider the set of fundamental memories $(\mathbf{x}^\xi, \mathbf{y}^\xi)$, for $\xi = 1, \ldots, 8$, presented in Table I. We stored these eight associations using the FAM models mentioned previously. The max–min FAM with threshold achieved perfect recall of all patterns. Junbo's FAM only succeeded in recalling the pairs $(\mathbf{x}^3, \mathbf{y}^3)$, $(\mathbf{x}^5, \mathbf{y}^5)$, $(\mathbf{x}^7, \mathbf{y}^7)$, and $(\mathbf{x}^8, \mathbf{y}^8)$. Kosko's FAM is only able to recall the pair $(\mathbf{x}^5, \mathbf{y}^5)$ prefectly while the generalized FAM of Chung-Lee using Lukasiewicz t-norm is unable to store and recall any of the patterns. From now on, we will refer to the latter model as the Lukasiewicz GFAM. We also stored the eight fundamental memories of Table I using the associative memory of Wang and Lu [29]. This model succeded in recalling the pairs $(\mathbf{x}^1, \mathbf{y}^1)$, $(\mathbf{x}^3, \mathbf{y}^3)$, $(\mathbf{x}^5, \mathbf{y}^5)$, and $(\mathbf{x}^8, \mathbf{y}^8)$.

Considering (39), we note that the threshold $\boldsymbol{\theta}$ that is used in a particular IFAM model does not depend on the t-norm. For the patterns $\mathbf{y}^\xi$ of Table I, we compute

$$\boldsymbol{\theta} = [0.5 \quad 0.6 \quad 0.3]^T. \tag{42}$$

Now, let us construct the weight matrices of some particular IFAM models using (39). We determine the following weight matrix of the Lukasiewics IFAM:

$$W_{Lk} = \begin{bmatrix} 0.7 & 1.0 & 1.0 & 0.9 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.5 & 0.8 & 0.8 & 0.7 & 1.0 \end{bmatrix}. \tag{43}$$

The recall of Lukasiewicz IFAM is performed using the Lukasiewicz t-norm within a max-t composition. In this case, $\left(W \circ \mathbf{x}^\xi\right) \vee \boldsymbol{\theta} = \mathbf{y}^\xi$ for all $\xi = 1, \ldots, 8$. The same is true for Goguen and Gödel IFAMs, whose synaptic weight matrices are given by

$$W_{Gg} = \begin{bmatrix} 0.6250 & 1.0000 & 1.0000 & 0.8571 & 1.0000 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.3750 & 0.6000 & 0.6667 & 0.5714 & 1.0000 \end{bmatrix} \tag{44}$$

and

$$W_{\text{Gd}} = \begin{bmatrix} 0.5 & 1.0 & 1.0 & 0.6 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.3 & 0.3 & 0.3 & 0.3 & 1.0 \end{bmatrix}. \tag{45}$$

TABLE II
ABSOLUTE STORAGE CAPACITY OF THE MODELS IN EXAMPLE 3.1

Model	Absolute storage capacity
Max-min FAM with threshold	8
Lukasiewicz IFAM	8
Gödel IFAM	8
Goguen IFAM	8
Junbo's FAM	4
AM of Wang and Lu	4
Kosko's FAM	1
Lukasiewicz GFAM	0

Recall that the Goguen IFAM employs max-product composition and the Gödel IFAM employs max-min composition during the recall phase. Table II summarizes the results concerning the absolute storage capacity of the models presented in this example.

Now, let us investigate the relationship between the Gödel IFAM and Liu's max–min FAM with threshold. Upon presentation of an input pattern $\mathbf{x}$, the latter model computes the following output pattern $\mathbf{y}$:

$$\mathbf{y} = [W \circ (\mathbf{x} \vee \mathbf{c})] \vee \mathbf{d} \tag{46}$$

where $W \in [0,1]^{m \times n}$ is the synaptic weight matrix and $\mathbf{c} \in [0,1]^n$ and $\mathbf{d} \in [0,1]^m$ are threshold vectors. The synaptic weight matrix W and the threshold $\mathbf{d}$ are determined using (39) with Gödel R-implication. The threshold vector $\mathbf{c} = [c_1, c_2, \ldots, c_n]^T$ is given by

$$c_i = \begin{cases} \bigwedge_{j \in D_i} \bigwedge_{k \in LE} y_j^\xi, & \text{if } D_i \neq \emptyset \\ 0, & \text{if } D_i = \emptyset \end{cases} \tag{47}$$

where $LE_{ij} = \{\xi \in \{1, 2, \ldots, p\} \mid x_i^\xi \leq y_j^\xi\}$ and $D_i = \{j \in \{1, 2, \ldots, m\} \mid LE_{ij} \neq \emptyset\}$. Thus, the difference between the Gödel IFAM and the max-min FAM with threshold lies in the threshold $\mathbf{c}$. However, due to the monotonicity of the max–min composition (9), the following equations hold true:

$$\begin{aligned} \mathbf{y} &= [W \circ (\mathbf{x} \vee \mathbf{c})] \vee \mathbf{d} & (48) \\ &= [(W \circ \mathbf{x}) \vee (W \circ \mathbf{c})] \vee \mathbf{d} & (49) \\ &= (W \circ \mathbf{x}) \vee \mathbf{f} & (50) \end{aligned}$$

where $\mathbf{f} = (W \circ \mathbf{c}) \vee \mathbf{d}$. These equalities show that the action of both thresholds $\mathbf{c}$ and $\mathbf{d}$ can be captured in terms of the new threshold $\mathbf{f}$. Moreover, Corollary 3.1 states that if there exists a threshold $\mathbf{f}$ such that all input–output pairs are successfully stored in the max-min FAM with threshold then the Gödel

TABLE III
SET OF INPUT AND OUTPUT PAIRS USED IN THE FORECASTING APPLICATION

ξ	$\mathbf{x}^\xi$	$\mathbf{y}^\xi$
1	$[1.0, 0.5, 0, 0, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
2	$[0.5, 1.0, 0.5, 0, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
3	$[0.5, 1.0, 0.5, 0, 0]^T$	$[0, 0.5, 1.0, 0.5, 0]^T$
4	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
5	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0, 0.5, 1.0, 0.5, 0]^T$
6	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0, 0, 0.5, 1.0, 0.5]^T$
7	$[0, 0, 0.5, 1.0, 0.5]^T$	$[0, 0, 0.5, 1.0, 0.5]^T$
8	$[0, 0, 0.5, 1.0, 0.5]^T$	$[0, 0, 0, 0.5, 1.0]^T$
9	$[0, 0, 0, 0.5, 1.0]^T$	$[0, 0, 0, 0.5, 1.0]^T$

IFAM can also store all input–output pairs and $\boldsymbol{\theta} \geq \mathbf{f}$. In this respect, the performance of the Gödel IFAM is at least as good as the performance of the max–min FAM with threshold. In view of this fact, we can view the IFAM as a generalization of the max–min FAM with threshold for any $\max\text{-}t$ composition. In Section III-A, we will present an experiment where the Lukasiewicz IFAM outperformed Liu's model.

A. Application of IFAMs as Fuzzy Rule-Based Systems

Fuzzy associative memories such as the IFAM can be used to implement mappings of fuzzy rules. In this case, a set of rules in the form of human-like IF–THEN conditional statements are stored. In this subsection, we briefly present an application of the IFAM model to the problem of forecasting manpower. Specifically, we consider a problem presented in [30]. This problem consists of assessing the engineering manpower requirement in steel manufacturing industry in the state of West Bengal, India. Initially, we have five linguistic values A_i, $i = 1, \ldots, 5$ and a set of fuzzy conditional statements such as "If the manpower of year n is A_i, then that of year $n+1$ is A_j". Hence, we obtain the set of input–output pairs given by Table III and we stored them in an IFAM model using (39). For instance, synthesizing the synaptic weight matrix using the Lukasiewicz implication yields

$$W_{Lk} = \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 \\ 1.0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1.0 \end{bmatrix}. \quad (51)$$

Note that we have $\boldsymbol{\theta} = [0, 0, 0, 0, 0]^T$. If W is the synaptic weight matrix and $\boldsymbol{\theta}$ is the threshold obtained after the learning process, then the predicted manpower of year $n+1$ is given by the following equation:

$$A_{n+1} = (W \circ A_n) \vee \boldsymbol{\theta} \quad (52)$$

where A_n is the manpower of year n and $\circ$ is the $\max\text{-}t$ composition.

Using Kosko's FAM, Choudhury et. al. found an average error of 2.669% whereas the statistical methods ARIMA1 and ARIMA2 exhibited average errors of 9.79% and 5.48%. Following the same procedure using the Lukasiewicz generalized FAM of Chung and Lee, the max–min FAM with threshold of Liu, the associative memory of Wang and Lu, the Gödel IFAM, the Goguen IFAM and, the Lukasiewicz IFAM, we obtained the average errors presented in Table IV. Kosko's FAM and the Lukasiewicz GFAM produced the same results. The associative memory of Wang and Lu and the Goguen IFAM also produced the same results. In addition, the max–min FAM with threshold and the Gödel IFAM also produced the same results. Fig. 1 plots the manpower data of the years 1984 through 1995. The actual values are compared to the predictions obtained by some of these methods.

TABLE IV
AVERAGE ERRORS IN FORECASTING MANPOWER

Method	Average Error
Lukasiewcz IFAM	2.29%
Kosko's FAM	2.67%
Lukasiewicz GFAM	2.67%
Gödel IFAM	2.73%
Max-min FAM with threshold	2.73%
Goguen IFAM	2.99%
AM of Wang and Lu	2.99%
ARIMA2	5.48%
ARIMA1	9.79%

The Lukasiewicz IFAM outperformed Kosko's FAM, the Lukasiewicz generalized FAM, the max-min FAM with threshold, and the statistical method ARIMA with respect to the purpose of forecasting [30]. This experiment indicates the utility of the IFAM model as a fuzzy rule-based system.

IV. BASIC PROPERTIES OF AUTOASSOCIATIVE FUZZY IMPLICATIVE MEMORIES

Let $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p] \in [0,1]^{n \times p}$ be the matrix whose columns are the original patterns. In this section, we focus on AFIMs whose weight matrices are synthesized in terms of a $\circledast_R$-product of X and X^T. Thus, we have

$$W = X \circledast_R X^T \text{ and } \boldsymbol{\theta} = \bigwedge_{\xi=1}^{p} \mathbf{x}^\xi. \quad (53)$$

Upon presentation of an input pattern $\tilde{\mathbf{x}}$, the AFIM W generates the pattern

$$\mathbf{x} = (W \circ \tilde{\mathbf{x}}) \vee \boldsymbol{\theta} \quad (54)$$

as an output pattern.

A. Convergence

Suppose that the AFIM W is employed with feedback. In other words, given an input pattern $\mathbf{x}(0) = \tilde{\mathbf{x}}$, we iterate the following equation for $k = 0, 1, \ldots$ until convergence:

$$\mathbf{x}(k+1) = (W \circ \mathbf{x}(k)) \vee \boldsymbol{\theta}. \quad (55)$$

The following equations reveal that $\mathbf{x}(k+1) = \mathbf{x}(k)$ for all $k \geq 1$:

$$\begin{aligned} \mathbf{x}(k+1) &= (W \circ \mathbf{x}(k)) \vee \boldsymbol{\theta} && (56) \\ &= (W \circ ((W \circ \mathbf{x}(k-1)) \vee \boldsymbol{\theta})) \vee \boldsymbol{\theta} && (57) \\ &= ((W \circ W) \circ \mathbf{x}(k-1)) \vee (W \circ \boldsymbol{\theta}) \vee \boldsymbol{\theta} && (58) \\ &= (W \circ \mathbf{x}(k-1)) \vee \boldsymbol{\theta} = \mathbf{x}(k) && (59) \end{aligned}$$

for all $k \geq 1$. Here, we used the associativity and monotonicity (or distributivity over the maximum) of $\max\text{-}t$ composition and

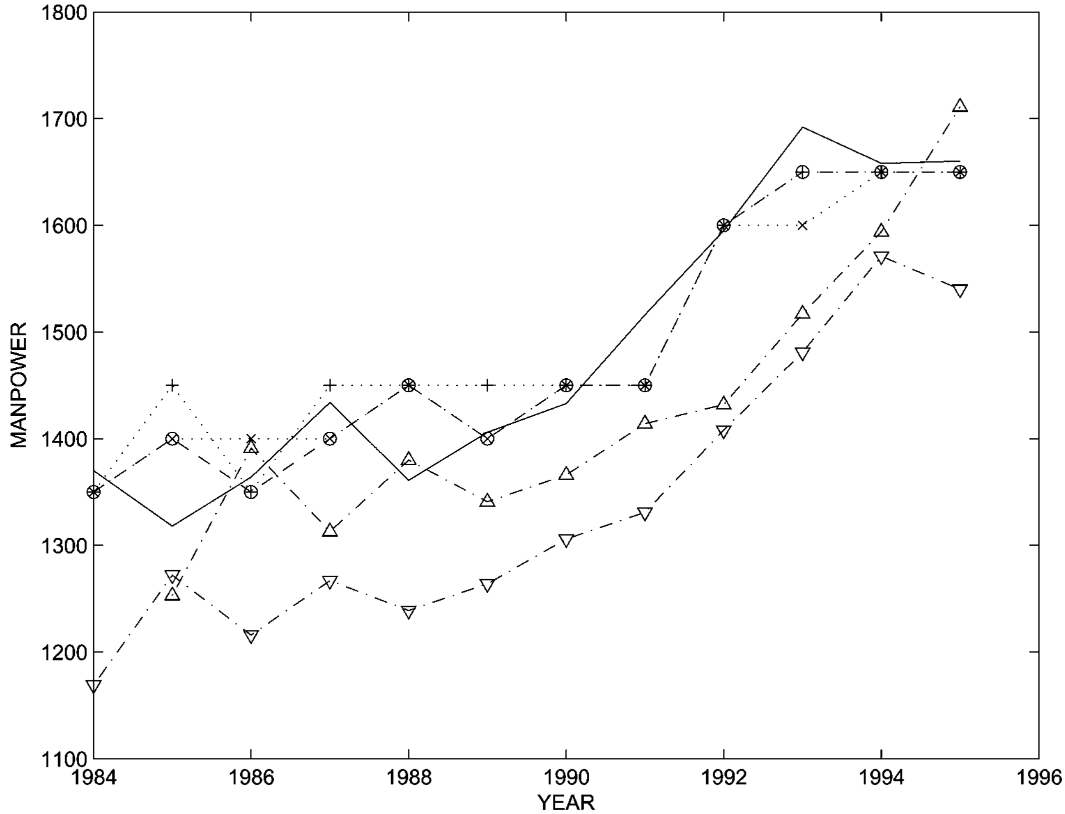


Fig. 1. Predictions of manpower. The continuous line represents the actual manpower. The dashed line marked by "o" corresponds to the Lukasiewicz IFAM model, the dotted line marked by "×" corresponds to Kosko's FAM model and the Lukasiewicz generalized FAM, and the dotted line marked by "+" corresponds to the max–min FAM with threshold and Gödel IFAM. The lines marked by "△" and "▽" represent ARIMA1 and ARIMA2.

the fact that W is reflexive, $\max\text{-}t$ idempotent and $W \circ \boldsymbol{\theta} = \boldsymbol{\theta}$. The first two facts are proven in Theorem 4.1 and the last one is a consequence of Theorem 4.3.

Theorem 4.1: The matrix $W = X \circledast_R X^T$ is reflexive and $\max\text{-}t$ idempotent. In other words, we have $I \leq W$ and $W \circ W = W$, where $\circ$ is the $\max\text{-}t$ composition with the t-norm used in R-implicative fuzzy learning.

We conclude that implicative fuzzy autoassociative memories exhibit one-pass convergence and thus there is no need for feedback to recall a pattern. Furthermore, due to the reflexivity of W and the monotonicity of the $\max\text{-}t$ composition, the input pattern is always less than or equal to the retrieved pattern, i.e., $W \circ \mathbf{x} \geq \mathbf{x}$ for all $\mathbf{x} \in [0,1]^n$.

B. Storage Capacity

This subsection answers the question as to how many patterns can be stored in an AFIM. Let $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$ be a set of p patterns, our task consists in determining the maximum number of patterns $\mathbf{x}^\xi$ such that $(W \circ \mathbf{x}^\xi) \vee \boldsymbol{\theta} = \mathbf{x}^\xi$. First of all, note that the threshold $\boldsymbol{\theta}$ does not affect the storage capacity of an AFIM because $\boldsymbol{\theta}$ constitutes the minimum of the stored patterns and W is reflexive. Thus, it suffices to determine the maximum number of patterns p such that $W \circ \mathbf{x}^\xi = \mathbf{x}^\xi$ holds for all $\xi = 1, \ldots, p$. We begin by presenting the following theorem.

Theorem 4.2: Let $X \in [0,1]^{n \times p}$. If $W = X \circledast_R X^T$, then $W \circ X = X$.

Proof: In view of Theorem 2.1 with $Y = X$, it suffices to show that $X \leq W \circ X$. This inequality arises immediately from a combination of the facts that W is reflexive and that the $\max\text{-}t$ composition is monotonic: $X = I \circ X \leq W \circ X$. ■

Note that Theorem 4.2 imposes no restrictions on the size of X. Therefore, as many patterns as desired can be stored. In particular, if the patterns are binary, i.e., $\mathbf{x}^\xi \in \{0,1\}^n$, then 2^n patterns can be stored. Hence, the absolute storage capacity of implicative fuzzy autoassociative memory exceeds the capacity of the Hopfield net of $n/(4 \log n)$ random binary patterns [38].

Finishing our discusion about the storage capacity of implicative fuzzy autoassociative memories, we need to point out that, using fuzzy implicative learning, the weight matrix W converges rapidly to the identity matrix I for $p \to \infty$. Thus, the error correction capability decreases considerably as more and more patterns are stored in the memory. Fig. 2 plots the Frobenius distances $||W - I||_F$ for weight matrices $W \in [0,1]^{100 \times 100}$ as a function of p, the percentage of stored patterns. This experiment was performed by taking the mean after 1000 simulations. The line marked by "+" corresponds to the weight matrix of Lukasiewicz IFAM. The line marked by "△" refers to the weight matrix of Gödel IFAM. The lines marked by "×" and "o" represent the Goguen IFAM and the autoassociative memory of Wang and Lu, respectively. Let W_{XX}^{prod} denote the weight-matrix of Wang and Lu's autoassociative memory. We would like to draw the reader's attention to the fact that $||W_{\mathbf{G}g} - I||_F$ equals $||W_{XX}^{prod} - I||_F$ in this experiment for a large fundamental memory set. We will provide an explanation for this fact in Section V.

The Lukasiewicz implicative learning recipe exhibits the slowest convergence to the identity matrix among these four

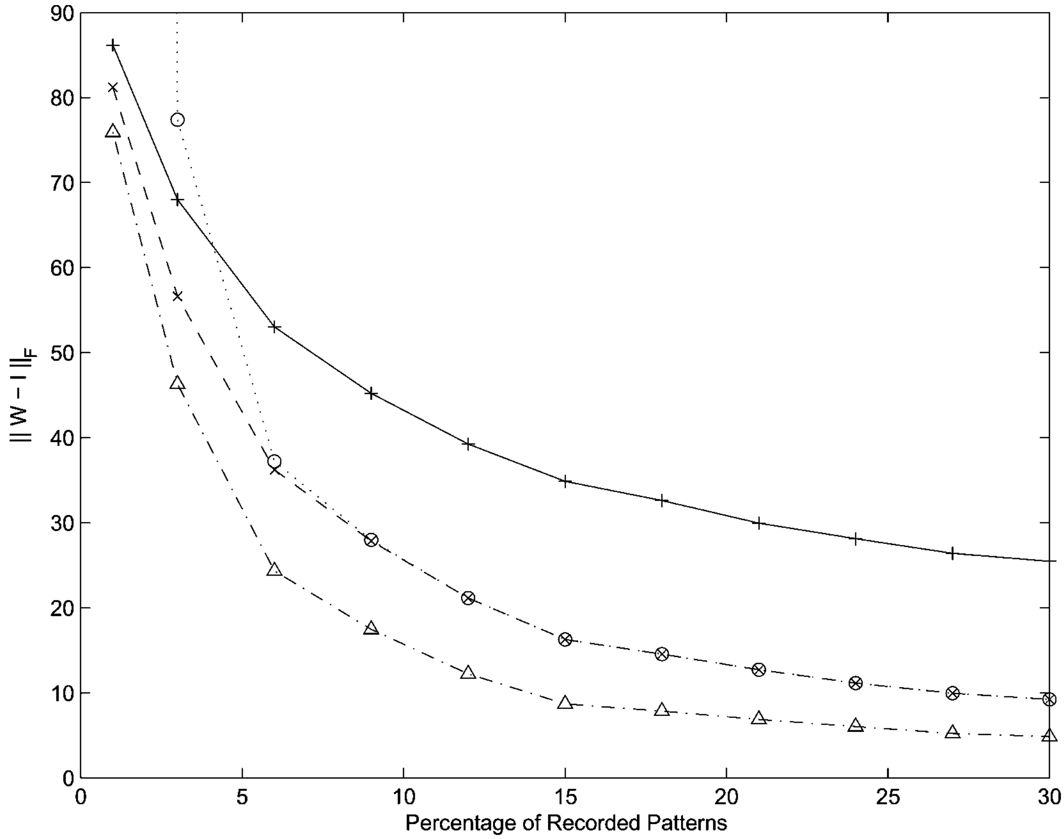


Fig. 2. Graph of $\|W-I\|_F$ versus the percentage of recorded patterns. The line marked by "+" corresponds to the weight matrix of Lukasiewicz IFAM. The line marked by "△" refers to the weight matrix of Gödel IFAM. The lines marked by "×" and "o" represent the Goguen IFAM and the associative memory of Wang and Lu, respectively.

learning recipes. Thus, we expect a better tolerance with respect to noise of the Lukasiewicz IFAM in comparison with the Gödel IFAM, the Goguen IFAM, and the associative memory of Wang and Lu. Despite the fact that determining the best t-norm for a given problem is an open problem, this observation motivates our choice of the Lukasiewicz IFAM in the forecasting problem in Section III-A and in Example 4.1.

C. *Tolerance With Respect to Noisy Patterns*

Since a certain error correction capability is a desired property for every associative memory model, we will discuss the ability of implicative fuzzy autoassociative memories to deal with incomplete or noisy patterns. For simplicity, we introduce the following terminology [21]: A pattern $\tilde{\mathbf{x}}$ is called an *eroded* version of a pattern $\mathbf{x}$ if and only if $\tilde{\mathbf{x}} \leq \mathbf{x}$. Similarly, a pattern $\tilde{\mathbf{x}}$ is called a *dilated* version of $\mathbf{x}$ if and only if $\tilde{\mathbf{x}} \geq \mathbf{x}$. We will show that the AFIM model exhibits tolerance with respect to eroded patterns. Similarly, the dual AFIM model exhibits tolerance with respect to dilated patterns.

Theorem 4.3: Let $X \in [0,1]^{n\times p}$ and $W = X \circledast_R X^T$. For every input pattern $\mathbf{x} \in [0,1]^n$, the output $(W \circ \mathbf{x}) \vee \boldsymbol{\theta}$ of the AFIM W is the *supremum* ("sup") of $\mathbf{x}$ in the set of fixed points of W greater than $\boldsymbol{\theta}$. In other words, $(W \circ \mathbf{x}) \vee \boldsymbol{\theta}$ is the smallest fixed point $\mathbf{y}$ of W such that $\mathbf{y} \geq \mathbf{x}$ and $\mathbf{y} \geq \boldsymbol{\theta}$.

Proof: Let $\mathbf{x} \in [0,1]^n$ be arbitrary. On one hand, the pattern $(W \circ \mathbf{x}) \vee \boldsymbol{\theta}$ represents an upper bound of $\mathbf{x}$ and $\boldsymbol{\theta}$ because $(W \circ \mathbf{x}) \vee \boldsymbol{\theta} \geq (I \circ \mathbf{x}) \vee \boldsymbol{\theta} \geq \mathbf{x} \vee \boldsymbol{\theta}$ by Theorem 4.1 and by monotonicity of the max-t composition. On the other hand, the fixed point $(W \circ \mathbf{x}) \vee \boldsymbol{\theta}$ is the least upper bound in the set of fixed points of AFIM since every fixed point $\mathbf{y} \geq \mathbf{x}$ satisfies $\mathbf{y} = (W \circ \mathbf{y}) \vee \boldsymbol{\theta} \geq (W \circ \mathbf{x}) \vee \boldsymbol{\theta}$. ■

Fig. 3. Patterns of Example 4.1.

1) Example 4.1: Consider the 12 patterns shown in Fig. 3. These are gray-scale images $\mathbf{x}^\xi \in [0,1]^{56\times 46}$, $\xi = 1,\ldots,12$ from the faces database of AT&T Laboratories, Cambridge, MA [39]. This database contains files in PGM format, and the size of each image is 92 × 112 pixels, with 256 gray levels per pixel.

Fig. 4. Patterns with erosive noise used as input in Example 4.1.

Fig. 5. Patterns retrieved by Lukasiewicz AFIM when the input are the patterns of Fig. 4.

Fig. 6. Patterns retrieved by Kosko's FAM, Lukasiewicz generalized FAM, the Gödel AFIM, and the Goguen AFIM, respectively, when the input is the first pattern of Fig. 4.

We downsized the original images using neighbor interpolation. In order to verify the tolerance of the AFIM model with respect to noise, we define $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{12}] \in [0,1]^{2576\times 12}$, where each $\mathbf{x}^\xi$ is a vector obtained using the standard row-scan method. Then, we treated each pattern as a fuzzy set and we constructed the matrix $W = X \circledast_R X^T$ using the Lukasiewiscz implication given by (21). Fig. 4 displays eroded versions $\tilde{\mathbf{x}}^1, \tilde{\mathbf{x}}^2, \ldots, \tilde{\mathbf{x}}^{12}$ of the fundamental memories which we presented as inputs to the AFIM W. These patterns contain erosive noise that was generated by subtracting the negative part of a Gaussian distribution with zero mean and variance 0.2. Fig. 5 shows the patterns $W \circ \tilde{\mathbf{x}}^\xi$, where $\xi = 1, \ldots, 12$, that were retrieved by the Lukasiewicz AFIM. The Lukasiewicz AFIM succeeded in recalling the original patterns almost perfectly. We also conducted the same experiment using Kosko's FAM, the Lukasiewicz generalized FAM, the Gödel AFIM, and the Goguen AFIM. Fig. 6 shows the pattern $W \circ \tilde{\mathbf{x}}^1$ recalled by these associative memory models. Note that Kosko's FAM and the Lukasiewicz generalized FAM fail to demonstrate an adequate performance on this task due to a high amount of crosstalk between the stored patterns.

Fig. 7 plots the normalized mean square errors (NMSEs)

$$\frac{\|\mathbf{x}^1 - W \circ \tilde{\mathbf{x}}^1\|_2}{\|\mathbf{x}^1\|_2} \tag{60}$$

that was generated by various fuzzy associative memory models for different variances of Gaussian noise whose negative part was subtracted from $\mathbf{x}^1$. This experiment was performed taking the mean after 100 simulations. We also conducted the experiment using Kohonen's OLAM [13], [15]. The Lukasiewicz AFIM yielded the best performance in this experiment. Note that both the Kosko FAM and the Lukasiewicz GFAM showed a NMSE greater than zero even for uncorrupted input patterns, i.e., when the noise variance is zero. Thus, these two fuzzy associative memories models fail to store the pattern $\mathbf{x}^1$ in contrast to the AFIM models and the OLAM that succeeded in storing the complete fundamental memory set. Fig. 7 also reveals that the NMSEs of the AFIM models are almost constant. These models will produce almost the same outputs even if the variance of Gaussian noise increases from 0.1 to 0.3. Fig. 8 shows the output patterns of the Lukasiewicz AFIM for inputs that were corrupted using the negative parts of Gaussian noise with variances 0.01, 0.05, 0.1, and 0.5. Note that all outputs are very similar. We also performed the latter experiment using the complete face database, i.e., we stored 40 patterns using the Lukasiewicz AFIM and the OLAM. Table V compares the NMSEs for the cases where 12 patterns are stored and where 40 patterns are stored in a Lukasiewicz AFIM and in a OLAM model. The NMSE was obtained taking the mean of 1000 simulation for inputs corrupted using the negative part of a gaussian distribution with variance 0.01, 0.05, 0.1, and 0.5. Note that the Lukasiewicz AFIM outperformed the OLAM model in both cases.

Recall that Theorem 4.3 relates the retrieved pattern for a given input to the set of fixed points of an AFIM. Let us examine the fixed points of AFIMs.

D. Fixed Points

First, we will present results about the fixed points of the matrix $W = X \circledast_R X^T$ used in the AFIM model. Then we will consider the fixed points of the AFIM. Recall that $\mathbf{y} \in [0,1]^n$ is a fixed point of W if and only if $W \circ \mathbf{y} = \mathbf{y}$ and $\mathbf{y}$ is a fixed point of the AFIM if and only if $(W \circ \mathbf{y}) \vee \boldsymbol{\theta} = \mathbf{y}$.

Theorem 4.4: Let $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p] \in [0,1]^{n\times p}$, $W = X \circledast_R X^T$. A pattern $\mathbf{y}$ is a fixed point of W if $\mathbf{y} = \mathbf{c}$ for some

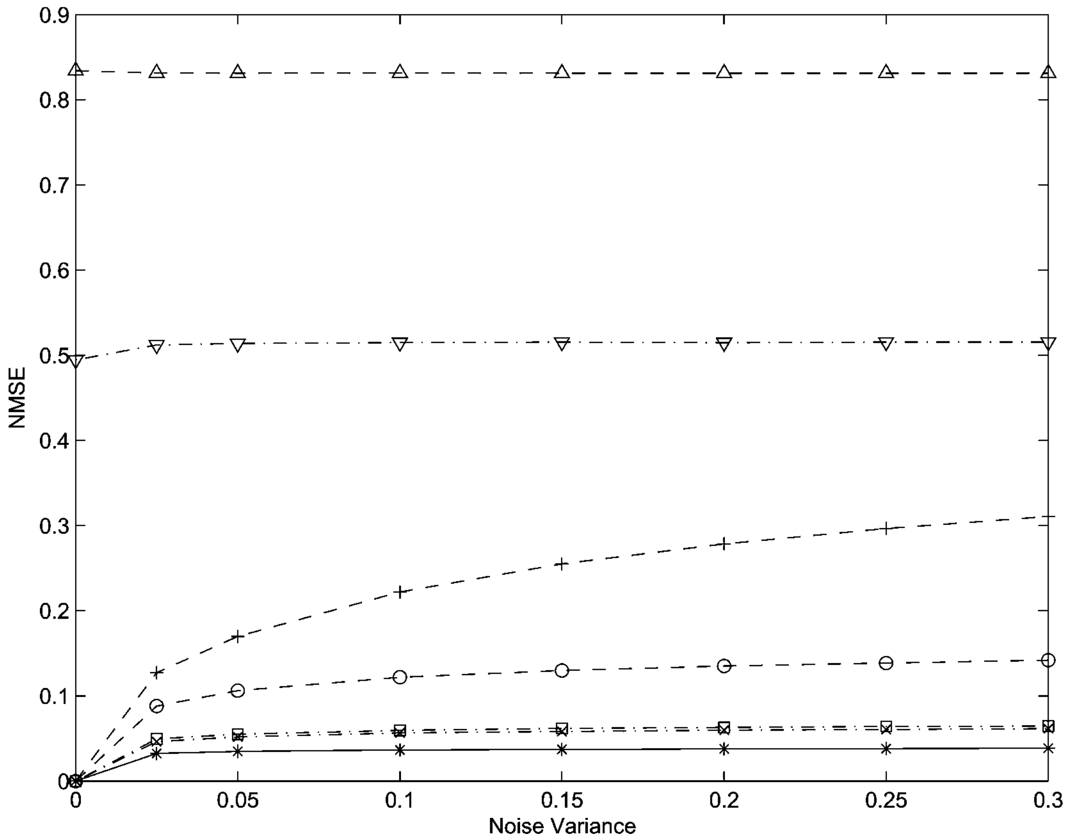


Fig. 7. The normalized mean square error versus the variance of the gaussian noise whose negative part was subtracted to $\mathbf{x}^1$. The lines marked by the symbols "$*$," "o," "□," "△," "▽," and "×" represent the NMSE of the Lukasiewicz AFIM, Gödel AFIM, and Goguen AFIM, the Kosko's FAM, the Lukasiewicz GFAM, and the AM of Wang and Lu, respectively. The line marked by the symbol "+" corresponds to the OLAM model.

Fig. 8. At the top, eroded patterns $\tilde{\mathbf{x}}^1$ computed using the negative part of gaussian noise with variance 0.01, 0.05, 0.1, and 0.5. At the botton, the respective recalled pattern by Lukasiewicz AFIM.

TABLE V
THE NMSE'S OF LUKASIEWICZ AFIM AND OLAM MODEL STORING 12 AND 40 PATTERNS WHERE THE INPUTS WERE CORRUPTED USING THE NEGATIVE PART OF A GAUSSIAN DISTRIBUTION WITH VARIANCE 0.01, 0.05, 0.1, AND 0.5

NMSE of Lukasiewicz AFIM	Variance			
	0.01	0.05	0.1	0.5
12 patterns stored	0.0283	0.0347	0.0365	0.0391
40 patterns stored	0.0580	0.0780	0.0837	0.0919
NMSE of OLAM	**Variance**			
	0.01	0.05	0.1	0.5
12 patterns stored	0.0853	0.1700	0.2220	0.3473
40 patterns stored	0.0869	0.1721	0.2247	0.3492

constant vector $\mathbf{c} = [c, c, \ldots, c]^T \in [0,1]^n$ or if $\mathbf{y}$ is of the following form:

$$\mathbf{y} = \bigvee_{l=1}^{k} \bigwedge_{\xi \in L_l} \mathbf{x}^\xi \tag{61}$$

for some $L_l \subseteq \{1, \ldots, p\}$ and some $k \in \mathbb{N}$.

Proof: Let $i \in \{1, 2, \ldots, n\}$ be an index of $\mathbf{x}$. Using the monotonicity of triangular norms, distributivity of $\vee$ and $\wedge$, and Theorem 4.2, we conclude that

$$(W \circ \mathbf{y})_i = \bigvee_{j=1}^{n} (w_{ij} \,\mathrm{t}\, y_j) \tag{62}$$

$$= \bigvee_{j=1}^{n} \left[w_{ij} \,\mathrm{t}\, \left(\bigvee_{l=1}^{k} \bigwedge_{\xi \in L_l} x_j^\xi \right) \right] \tag{63}$$

$$= \bigvee_{j=1}^{n} \left[\bigvee_{l=1}^{k} \bigwedge_{\xi \in L_l} \left(w_{ij} \,\mathrm{t}\, x_j^{\xi} \right) \right] \tag{64}$$

$$= \bigvee_{l=1}^{k} \bigwedge_{\xi \in L_l} \left[\bigvee_{j=1}^{n} \left(w_{ij} \,\mathrm{t}\, x_j^{\xi} \right) \right] \tag{65}$$

$$= \bigvee_{l=1}^{k} \bigwedge_{\xi \in L_l} x_i^{\xi} = y_i. \tag{66}$$

Since the diagonal elements of W are equal to 1, a constant pattern $\mathbf{c}$ remains fixed under an application of W due to the following facts:

$$(W \circ \mathbf{c})_i = \bigvee_{j=1}^{n} (w_{ij} \,\mathrm{t}\, c) = c \,\mathrm{t} \left(\bigvee_{j=1}^{n} w_{ij} \right) = c. \tag{67}$$

■

The expression $\bigvee_{l=1}^{k} \bigwedge_{\xi \in L_l} \mathbf{x}^{\xi}$, involving the symbols $\vee$, $\wedge$, and $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$, represents a lattice polynomial in $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$ [40]. Since the lattice $[0,1]^n$ is distributive, every lattice polynomial in $\mathbf{x}^1, \ldots, \mathbf{x}^p$ is of the form given by (61) and is a fixed point of W according to Theorem 4.4. Note that the threshold $\boldsymbol{\theta}$ is the smallest lattice polynomial in $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$ and, thus, $W \circ \boldsymbol{\theta} = \boldsymbol{\theta}$. Moreover, this threshold excludes the spurius states smaller than $\boldsymbol{\theta}$. Nevertheless, implicative fuzzy autoassociative memories have a large number of fixed points, that include the fundamental memories and many spurious states.

E. Binary Implicative Autoassociative Memories

Let us turn our attention to the binary autoassociative case, where $X = Y \in \{0,1\}^{n \times k}$. This case is of particular interest because many traditional associative memory models such as the Hopfield net, the BSB, and the ECAM are dealing with this case [4], [16].

First of all, note that in the binary case, all triangular norms, triangular co-norms and implications coincide. Consequently, only one IFAM model exists in this case. Thus, we may assume that—without loss of generality—$\mathrm{t} = \wedge$ (minimum), $\mathrm{s} = \vee$ (maximum), and $x \Rightarrow y = (1 - x) \vee y$.

In Section IV-B, we mentioned that autoassociative fuzzy implicative memories are endowed with unlimited storage capacity. In the binary autoassociative case, we can store 2^n patterns, where n is the length of the patterns. The general results on noise tolerance and fixed points of Sections IV-C and D also hold true for the binary case. However, there is a stronger result that characterizes the fixed points of binary autoassociative implicative memories. Before we enunciate this result, let us introduce the following notation: Given a set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p\}$ and a index $l \in \{1, 2, \ldots, n\}$, the symbol L_l denotes the subset of $\{1, 2, \ldots, p\}$ such that $x_l^{\xi} = 1$ for all $\xi \in L_l$. In other words, the set L_l consists of the indexes of the patterns whose lth entry equals 1.

Lemma 5.5: Let $\mathbf{e}^l$ be the lth column of the identity matrix. We have

$$W \circ \mathbf{e}^l = \bigwedge_{\xi \in L_l} \mathbf{x}^{\xi}. \tag{68}$$

The following theorem provides a precise characterization of the set of fixed points of a binary AFIM.

Theorem 4.6: Let $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p] \in \{0,1\}^{n \times p}$ and $W = X \circledast X^T$. A binary pattern $\mathbf{y}$ is a fixed point of W if and only if $\mathbf{y} = \mathbf{0}$ or $\mathbf{y} = \mathbf{1}$ or if $\mathbf{y}$ is a lattice polynomial in $\mathbf{x}^1, \ldots, \mathbf{x}^p$.

Proof: Theorem 4.4 implies that every lattice polynomial in $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$ and the constant pattern $\mathbf{y} = \mathbf{0}$ and $\mathbf{y} = \mathbf{1}$ are fixed points of W. Now, let $\mathbf{y}$ be a binary fixed point such that $\mathbf{y} \neq \mathbf{1}$ $\mathbf{y} \neq \mathbf{0}$. Evidently, the pattern $\mathbf{y}$ can be written in the form

$$\mathbf{y} = \bigvee_{l \in k} \mathbf{e}^l$$

, where k is the set of nonzero elements of $\mathbf{y}$. We are able to deduce the following equations for the ith element of the fixed point $\mathbf{y} = W \circ \mathbf{y}$, where $i = 1, \ldots, n$:

$$y_i = (W \circ \mathbf{y})_i = \bigvee_{j=1}^{n} (w_{ij} \wedge y_j) \tag{69}$$

$$= \bigvee_{j=1}^{n} \left(w_{ij} \wedge \bigvee_{l \in k} \delta_{lj} \right) = \bigvee_{l \in k} \left[\bigvee_{j=1}^{n} (w_{ij} \wedge \delta_{lj}) \right] \tag{70}$$

$$= \bigvee_{l \in k} \left(W \circ \mathbf{e}^l \right)_i = \bigvee_{l \in k} \left(\bigwedge_{\xi \in L_l} \mathbf{x}^{\xi} \right)_i. \tag{71}$$

■

Theorem 4.6 tells us that our application of the threshold

$$\theta = \bigwedge_{\xi \in 1}^{P} \mathbf{x}^{\xi}$$

in the binary AFIM possibly excludes fixed point $\mathbf{y} = \mathbf{0}$ in the binary autoassociative case. Moreover, together with Theorem 4.3, Theorem 4.6 shows that binary AFIMs and AMMs act in the same way provided that a certain weak condition on the stored patterns is satisfied [41], [42]. In Section V, we will show that these similarities between the IFAM and the MAM models are not restricted to the binary autoassociative case.

V. IFAM AND MORPHOLOGICAL ASSOCIATIVE MEMORIES

In this section, we will discuss the relationship between IFAMs and MAMs [21], [22].

Recall that the morphological associative memory W_{XY} calculates

$$\mathbf{y} = W_{XY} \boxed{\vee} \mathbf{x} \tag{72}$$

where the synaptic weight matrix is given by

$$W_{XY} = Y \boxed{\wedge} \left(-X^T \right). \tag{73}$$

Here, for matrices $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{k \times p}$, the matrices $C = A \boxed{\vee} B \in \mathbb{R}^{m \times n}$ and $D = A \boxed{\wedge} B \in \mathbb{R}^{m \times n}$ are defined by

$$c_{ij} = \bigvee_{l=1}^{p} (a_{il} + b_{lj}) \text{ and } d_{ij} = \bigwedge_{l=1}^{p} (a_{il} + b_{lj}) \tag{74}$$

for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The dual formulation is given by the equation $\mathbf{y} = M_{XY} \boxed{\wedge} \mathbf{x}$, where $M_{XY} = Y \boxed{\vee} \left(-X^T \right)$.

Let t and $\Rightarrow_R$ be the Lukasiewicz t-norm and the R-implication given by (15) and (21), respectively. Using this t-norm and R-implication, the elements of the synaptic weight matrix $W_{Lk} = (w_{ij}) \in [0,1]^{m\times n}$ will be

$$w_{ij} = \bigwedge_{\xi=1}^{p} \left(x_j^{\xi} \Rightarrow_R y_i^{\xi}\right) = \bigwedge_{\xi=1}^{p} \left[1 \wedge \left(1 - x_j^{\xi} + y_i^{\xi}\right)\right] \quad (75)$$

$$= 1 \wedge \left[1 + \bigwedge_{\xi=1}^{p} \left(y_i^{\xi} - x_j^{\xi}\right)\right] - 1 + 1 \quad (76)$$

$$= (1-1) \wedge \left[\bigwedge_{\xi=1}^{p} \left(y_i^{\xi} - x_j^{\xi}\right)\right] + 1 \quad (77)$$

$$= \left[\bigwedge_{\xi=1}^{p} \left(y_i^{\xi} - x_j^{\xi}\right)\right] \wedge 0 + 1. \quad (78)$$

Thus, the synaptic weight matrices of the Lukasiewicz IFAM and the MAM are related according to the equation $W_{Lk} = (W_{XY} \wedge 0) + 1$.

In the retrieval phase, the following pattern is recalled by the Lukasiewicz IFAM

$$(W_{Lk} \circ \mathbf{x})_i \vee \theta_i = \bigvee_{j=1}^{n} (w_{ij} \,\mathrm{t}\, x_j) \vee \theta_i \quad (79)$$

$$= \bigvee_{j=1}^{n} [0 \vee (w_{ij} + x_j - 1)] \vee \theta_i \quad (80)$$

$$= 0 \vee \left\{ \bigvee_{j=1}^{p} [(w_{ij} - 1) + x_j] \right\} \vee \theta_i \quad (81)$$

$$= \left\{ \bigvee_{j=1}^{n} \left[(W_{XY})_{ij} \wedge 0 + x_j\right] \right\} \vee \theta_i \quad (82)$$

for all $i = 1, \ldots, n$ since $\boldsymbol{\theta} \geq \mathbf{0}$. Thus, we obtain

$$(W_{Lk} \circ \mathbf{x}) \vee \boldsymbol{\theta} = [(W_{XY} \wedge 0) \boxvee \mathbf{x}] \vee \boldsymbol{\theta}. \quad (83)$$

We proceed by describing the relationship between the Lukasiewicz dual IFAM and the MAM M_{XY}. The entries of the synaptic weight matrix $M_{Lk} = (m_{ij}) \in [0,1]^{m\times n}$ of the dual IFAM can be computed as follows:

$$m_{ij} = \bigvee_{\xi=1}^{p} \left(x_j^{\xi} \Rrightarrow y_i\right) = \bigvee_{\xi=1}^{p} \left[0 \vee \left(y_i^{\xi} - x_j^{\xi}\right)\right] \quad (84)$$

$$= 0 \vee \left[\bigvee_{\xi=1}^{p} \left(y_i^{\xi} - x_j^{\xi}\right)\right]. \quad (85)$$

Thus

$$M_{Lk} = M_{XY} \vee 0. \quad (86)$$

In the retrieval phase, the following pattern is recalled by the Lukasiewicz dual IFAM:

$$(M_{Lk} \bullet \mathbf{x})_i \wedge \vartheta_i = \bigwedge_{j=1}^{n} (m_{ij} \,\mathrm{s}\, x_j) \wedge \vartheta_i \quad (87)$$

$$= \bigwedge_{j=1}^{n} [1 \wedge (m_{ij} + x_j)] \wedge \vartheta_i \quad (88)$$

$$= \left[\bigwedge_{j=1}^{n} (m_{ij} + x_j)\right] \wedge \vartheta_i \quad (89)$$

since $\vartheta_i \leq \mathbf{1}$. Thus, we obtain

$$(M_{Lk} \bullet \mathbf{x}) \wedge \boldsymbol{\vartheta} = (M_{Lk} \boxwedge \mathbf{x}) \wedge \boldsymbol{\vartheta} = [(M_{XY} \vee 0) \boxwedge \mathbf{x}] \wedge \boldsymbol{\vartheta}. \quad (90)$$

We summarize the results that we deduced in this section by stating the following theorem.

Theorem 5.1: Let $X \in [0,1]^{n\times p}$ and $Y \in [0,1]^{m\times p}$ be the input and output matrices, respectively. Let $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ denote the thresholds $\boldsymbol{\theta} = \bigwedge_{\xi=1}^{p} \mathbf{x}^{\xi}$ and $\boldsymbol{\vartheta} = \bigvee_{\xi=1}^{p} \mathbf{x}^{\xi}$. The Lukasiewicz IFAM and the MAM $W_{XY} = Y \boxwedge (-X)^T$ are related according to the following equations:

$$W_{Lk} = (W_{XY} \wedge 0) + 1 \quad (91)$$

$$(W_{Lk} \circ \mathbf{x}) \vee \boldsymbol{\theta} = [(W_{XY} \wedge 0) \boxvee \mathbf{x}] \vee \boldsymbol{\theta} \quad (92)$$

for all $\mathbf{x} \in [0,1]^n$.

Similarly, the Lukasiewicz dual IFAM and the MAM $M_{XY} = Y \boxvee (-X)^T$ are related according to the following equations:

$$M_{Lk} = M_{XY} \vee 0 \quad (93)$$

$$(M_{Lk} \bullet \mathbf{x}) \wedge \boldsymbol{\vartheta} = [(M_{XY} \vee 0) \boxwedge \mathbf{x}] \wedge \boldsymbol{\vartheta} \quad (94)$$

for all $\mathbf{x} \in [0,1]^n$.

Theorem 5.1 explains the relationship between the Lukasiewicz IFAM and the MAM model. Moreover, this theorem reveals that the learning phase as well as the retrieval phase of the Lukasiewicz IFAM can be described entirely in terms of the corresponding phases of the MAM model with threshold. Both models produce the same output when morphological operations are restricted to the interior of some hypercube. More precisely, the output patterns of the Lukasiewicz IFAM and the MAM W_{XY} coincide whenever $W_{XY} \in [-1,0]^{m\times n}$ and the threshold $\boldsymbol{\theta}$ is used in the recall phase of W_{XY}. Similarly, the output patterns of the Lukasiewicz dual IFAM and the MAM M_{XY} coincide whenever $M_{XY} \in [0,1]^{m\times n}$ and the threshold $\boldsymbol{\vartheta}$ is used in the recall phase of M_{XY}. Thus, if the fundamental memories are fuzzy and the synaptic weight matrix W_{XY} (M_{XY}, respectively) is restrained to the hypercube $[-1,0]^{m\times n}$ ($[0,1]^{m\times n}$, respectively), we can view the IFAM (the dual IFAM) as a generalization of the MAM model. The probability of this event rapidly approaches 1 as more and more patterns are stored. The synaptic weight matrix W_{XY} of the MAM model converges as fast as the synaptic weight matrix of Lukasiewicz IFAM in Fig. 2 to the identity matrix I for $p \to \infty$. An application of the MAM W_{XY} to the forecasting problem of Section III-A yields the same results as the ones predicted by the Lukasiewicz IFAM. We also stored the eight patterns of Example 3.1 in the MAM W_{XY} with and without threshold. The MAM W_{XY} with threshold achieved perfect recall of all patterns while the MAM W_{XY} without threshold only succeded in recalling the pairs $(\mathbf{x}^1,\mathbf{y}^1)$, $(\mathbf{x}^2,\mathbf{y}^2)$, $(\mathbf{x}^3,\mathbf{y}^3)$, $(\mathbf{x}^4,\mathbf{y}^4)$, $(\mathbf{x}^5,\mathbf{y}^5)$, and $(\mathbf{x}^8,\mathbf{y}^8)$.

Finally, we would like to point out a relationship between the Goguen IFAM and the associative memory model of Wang and Lu that is known as "*fuzzy morphological associative memory*" [29]. Apart from the previously defined matrix-products called *additive maximum* ("$\boxvee$") and *additive minimum* ("$\boxwedge$"), minimax algebra defines the matrix-products *multiplicative maximum* and *multiplicative minimum* (this terminology stems from image algebra that encompasses minimax algebra and linear algebra as subalgebras) [43]–[45]. For an $m \times k$ matrix A and an $k \times n$ matrix B with elements in $\mathbb{R}_{\infty}^{\geq 0} = \{r \in \mathbb{R} \cup +\infty : r \geq 0\}$, the *multiplicative maximum* $C = A \circledvee B$ and the *multiplicative minimum* $D = A \circledwedge B$ are the $m \times n$ matrices given by

$$c_{ij} = \bigvee_{l=1}^{k} (a_{il} \cdot b_{lj}) \text{ and } d_{ij} = \bigwedge_{l=1}^{k} (a_{il} \cdot' b_{lj}) \tag{95}$$

for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The operations $\cdot$ and $\cdot'$ differ as follows:

$$0 \cdot \infty = \infty \cdot 0 = 0 \tag{96}$$

$$0 \cdot' \infty = \infty \cdot' 0 = \infty \tag{97}$$

Otherwise, these products act as one would expect. Using $\circledwedge$ and $\circledvee$, we construct the synaptic weight matrix of a multiplicative minimum associative memory in terms of

$$W_{XY}^{\text{prod}} = Y \circledwedge X^* \tag{98}$$

where the i, jth element of X^* is $1/x_{ji}$. Here, we used the conventions $1/0 = \infty$ and $1/\infty = 0$. The recall phase is described by

$$\mathbf{y} = W_{XY}^{\text{prod}} \circledvee \mathbf{x}. \tag{99}$$

A multiplicative maximum associative memory M_{XY}^{prod} can be constructed in a similar fashion.

Note that the recall phases of the Goguen IFAM and W_{XY}^{prod} are identical—regardless of the fact whether a threshold is used or not. The difference between these two models lies in the recording phase because the following equations hold for all $i = 1, \ldots, m$ and for all $j = 1, \ldots, n$:

$$(W_{Gg})_{ij} = \bigwedge_{\xi=1}^{p} \left(x_j^\xi \Rightarrow y_i^\xi \right) \tag{100}$$

$$= \bigwedge_{\xi=1}^{p} \left(\frac{y_i^\xi}{x_j^\xi} \right) \wedge 1 = (W_{XY}^{prod})_{ij} \wedge 1. \tag{101}$$

Thus, the synaptic weight matrix W_{Gg} of Goguen IFAM is in $[0,1]^{m \times n}$ while the synaptic weight matrix W_{XY}^{prod} of multiplicative minimum model is in $(R_{\infty}^{\geq 0})^{m \times n}$. The Goguen dual IFAM and the multiplicative maximum model M_{XY}^{prod} differ in both storage and recall phases.

VI. Concluding Remarks

This paper introduces the IFAM, a neuro-fuzzy model described by Pedrycz logic neurons with thresholds where the synaptic weight matrices are computed using fuzzy implicative learning. A dual IFAM model arises from the duality relationships between maximum and minimum, triangular norm and co-norm, as well as implication and dual implication. Recall that every statement concerning the IFAM model proved in this paper corresponds to a dual statement concerning the dual IFAM model that arises by replacing the minimum with maximum, t-norm with s-norm, implication with dual implication, the products "$\circ$" with "$\bullet$," "$\circledast_R$" with "$\bar{\circledast}_R$," and *vice versa*. A particular IFAM model, such as the Lukasiewicz, Gödel, and Goguen IFAM, is associated with a certain R-implication.

Comparisons of these IFAM models with other FAM models by means of a simple example taken from [28] and a forecasting problem presented in [30] indicate the utility of IFAMs in applications as fuzzy rule-based systems. In the first experiment, the fundamental memory set was perfectly stored by the IFAM models under consideration and by Liu's FAM whereas other FAM models did not demonstrate perfect recall of the original patterns due to crosstalk. In the second experiment, concerning the prediction of manpower in steel industry, the Lukasiewicz IFAM outperformed the FAM models of Kosko, Liu, and Chung and Lee as well as the statistical methods ARIMA1 and ARIMA2. These experiments are dealing with applications of heteroassociative IFAM's.

In the autoassociative case, we speak of the AFIM. We were able to prove that—in sharp contrast to other FAM models—the AFIM exhibits unlimited storage capacity, one-pass convergence, and tolerance with respect to erosive noise. Similarly, the dual AFIM also exhibits unlimited storage capacity, one-pass convergence, and tolerance with respect to dilative noise. Thus, we recommend to consider the expected type of noise before selecting an AFIM model or a dual AFIM model. We illustrated the capabilities of AFIMs using face images from the database of AT&T Laboratories, Cambridge, MA. We noted that the functionality of an AFIM is completely determined by its fixed points and corresponding basins of attraction that are described in this paper. These results reveal that AFIMs and MAMs are almost identical in the binary case. Finally, we showed that the Lukasiewicz IFAM, which exhibited the best performance in our experiments, and the MAM model are very closely related if the fundamental memory set consists of fuzzy patterns. Further research is needed on how to choose the best IFAM or dual IFAM model for an arbitrary given application.

Appendix
Proofs of Theorems and Lemmas

a) Lemma 1.1: If $a^\xi, b^\xi \in [0,1]$, $\xi = 1, \ldots, p$ then the following inequality holds for every t-norm:

$$\bigwedge_{\xi=1}^{p} \left(a^\xi \,\mathrm{t}\, b^\xi\right) \geq \left(\bigwedge_{\xi=1}^{p} a^\xi\right) \,\mathrm{t}\, \left(\bigwedge_{\xi=1}^{p} b^\xi\right). \tag{102}$$

Proof: Consider arbitrary $a, b, c \in [0,1]$. By monotonicity of the t-norm, both $a \text{ t } c$ and $b \text{ t } c$ represent upper bounds of $(a \wedge b) \text{ t } c$. Thus, $(a \wedge b) \text{ t } c \le (a \text{ t } c) \wedge (b \text{ t } c)$ and

$$(a \wedge b) \text{ t } (c \wedge d) \le [a \text{ t } (c \wedge d)] \wedge [b \text{ t } (c \wedge d)] \quad (103)$$

$$\le (a \text{ t } c) \wedge (a \text{ t } d) \wedge (b \text{ t } c) \wedge (b \text{ t } d)$$

$$\le (a \text{ t } c) \wedge (b \text{ t } d). \quad (104)$$

Now, we will prove Lemma 1.1 by induction. It is trivial that (102) holds true for $p = 1$. Assume that (102) is satisfied for $p = k - 1 \ge 1$. Using this hypothesis and (104), we are able to conclude the proof of the lemma as follows:

$$\bigwedge_{\xi=1}^{k} \left(a^\xi \text{ t } b^\xi\right) = \left[\bigwedge_{\xi=1}^{k-1} \left(a^\xi \text{ t } b^\xi\right)\right] \wedge \left(a^k \text{ t } b^k\right) \quad (105)$$

$$\ge \left[\left(\bigwedge_{\xi=1}^{k-1} a^\xi\right) \text{ t } \left(\bigwedge_{\xi=1}^{k-1} b^\xi\right)\right] \wedge \left(a^k \text{ t } b^k\right) \quad (106)$$

$$\ge \left(\bigwedge_{\xi=1}^{k} a^\xi\right) \text{ t } \left(\bigwedge_{xi=1}^{k} b^\xi\right). \quad (107)$$

■

b) Lemma 1.2: Let $\text{t} : [0,1]^2 \rightarrow [0,1]$ be a continuous t-norm and let $\Rightarrow_R$: $[0,1]^2 \longrightarrow [0,1]$ be the corresponding R-implication. For all $a, b, c \in [0,1]$, we have

$$(a \Rightarrow_R b) \text{ t } (b \Rightarrow_R c) \le (a \Rightarrow_R c). \quad (108)$$

Proof: Let $x^* = (a \Rightarrow_R b)$, $y^* = (b \Rightarrow_R c)$ and $z^* = (a \Rightarrow_R c)$. Let z denote $x^* \text{ t } y^*$. It suffices to show that $z \le z^*$.

The commutativity and monotonicity of t-norm and the fact that $bt\,(b \Rightarrow_R c) \le c$ imply that

$$a \text{ t } z = a \text{ t } (x^* \text{ t } y^*) = (a \text{ t } x^*) \text{ t } y^* \le b \text{ t } y^* \le c \quad (109)$$

and, thus, the inequality $a \text{ t } z \le c$ holds. Since $z^* = (a \Rightarrow_R c) = \sup_z\{a \text{ t } z \le c\}$ by the definition of the R-implication $\Rightarrow_R$, we have $z \le z^*$ for all z such that $a \text{ t } z \le c$. Therefore, this inequality holds in particular for $z = x^* \text{t} y^*$. ■

Lemma 1.2 is a weak version of the syllogism of classical logic where the inequality symbol is replaced by an equality symbol. Given the results of Lemmas 1.1 and 1.2, we are now able to show that W is t-transitive.

c) Lemma 1.3: The matrix $W = X \circledast_R X^T$ is t-transitive. In other words, we have $W \circ W \le W$, where $\circ$ is the $\max$-tth composition with the t-norm used in R-implicative fuzzy learning.

Proof: To begin with, let us apply Lemma 1.2 to triples of entries of $X \in [0,1]^{n \times k}$ We obtain

$$\left(x_j^\xi \Rightarrow_R x_l^\xi\right) \text{ t } \left(x_l^\xi \Rightarrow_R x_i^\xi\right) \le \left(x_j^\xi \Rightarrow_R x_i^\xi\right) \quad (110)$$

for all $\xi = 1, 2, \ldots, p$ and for all $i, j, l = 1, 2, \ldots, n$. Applying the principle of closing of minimax algebra [43] to (110) yields

$$\bigwedge_{\xi=1}^{p} \left[\left(x_j^\xi \Rightarrow_R x_l^\xi\right) \text{ t } \left(x_l^\xi \Rightarrow_R x_i^\xi\right)\right] \le \bigwedge_{\xi=1}^{p} \left(x_j^\xi \Rightarrow_R x_i^\xi\right) \quad (111)$$

for all $\xi = 1, \ldots, p$ and for all $i, j, l = 1, \ldots, n$. Using this inequality and Lemma 1.1, we are able to deduce the following inequalities for all $l = 1, \ldots, n$:

$$w_{ij} = \bigwedge_{\xi=1}^{p} \left(x_j^\xi \Rightarrow_R x_i^\xi\right) \quad (112)$$

$$\ge \bigwedge_{\xi=1}^{p} \left[\left(x_j^\xi \Rightarrow_R x_l^\xi\right) \text{ t } \left(x_l^\xi \Rightarrow_R x_i^\xi\right)\right] \quad (113)$$

$$\ge \left[\bigwedge_{\xi=1}^{p} \left(x_j^\xi \Rightarrow_R x_l^\xi\right)\right] + \text{t} \left[\bigwedge_{\xi=1}^{p} \left(x_l^\xi \Rightarrow_R x_i^\xi\right)\right] \quad (114)$$

$$= w_{lj} \text{ t } w_{il} \quad (115)$$

for all $i, j = 1, \ldots, n$. Thus, we obtain $\bigvee_{l=1}^{n} (w_{il} \text{ t } w_{lj}) \le w_{ij}$ for all $i, j, l1, \ldots, n$ which implies and W is t-transitive. ■

Proof of Theorem 4.1: We begin by showing that W is reflexive. Note that

$$w_{ii} = \bigwedge_{\xi=1}^{p} \left(x_i^\xi \Rightarrow_R x_i^\xi\right) = \bigwedge_{\xi=1}^{p} 1 = 1. \quad (116)$$

Now, assume that W is reflexive and t-transitive. We will prove that W is $\max$-t idempotent. Note that we have the following inequalities in view of the t-transitivity of W and the definition of max-t composition:

$$w_{ij} \ge (W \circ W)_{ij} = \bigvee_{l=1}^{n} (w_{il} \text{ t } w_{lj}) \quad (117)$$

for all $i, j = 1, \ldots, n$. Applying the principle of opening [43] to the latter expression yields

$$\bigvee_{l=1}^{n} (w_{il} \text{ t } w_{lj}) \ge w_{ii} \text{ t } w_{ij} \quad (118)$$

which equals $1 \text{ t } w_{ij} = w_{ij}$ due to the reflexivity of W and the t-norm boundary property. Combining these facts with (117) implies that $W \circ W$ is bounded both from above and from below by W which concludes the proof of the equation $W \circ W = W$. ■

Proof of Lemma 4.5: The proof employs the definitions of W and L_l as well as some properties of the operations "$\wedge$," "$\vee$," and "$\Rightarrow$." Note that $\mathbf{e}^l(j)$ equals δ_{jl}, where $\delta_{jl} = 1$, if $j = l$ and 0, otherwise. For all $i, l = 1, 2, \ldots, n$, we obtain

$$\left(W \circ \mathbf{e}^l\right)_i = w_{il} = \bigwedge_{\xi=1}^{p} \left(x_l^\xi \Rightarrow x_i^\xi\right)$$

$$= \bigwedge_{\xi=1}^{p} (1 - x_l^\xi) \vee x_i^\xi$$

$$= \left[\bigwedge_{\xi \in L_l} \left(1 - x_l^\xi\right) \vee x_i^\xi\right] \wedge \left[\bigwedge_{\xi \notin L_l} \left(1 - x_l^\xi\right) \vee x_i^\xi\right]$$

$$= \left[\bigwedge_{\xi \in L_l} (1 - 1) \vee x_i^\xi\right] \wedge \left[\bigwedge_{\xi \notin L_l} (1 - 0) \vee x_i^\xi\right]$$

$$= \left[\bigwedge_{\xi \in L_l} x_i^\xi\right] \wedge \left[\bigwedge_{\xi \notin L_l} 1\right] = \bigwedge_{\xi \in L_l} x_i^\xi.$$

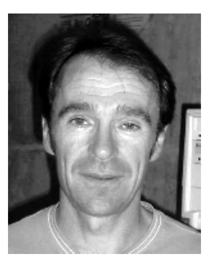
■

ACKNOWLEDGMENT

The authors would like to thank Dr. F. Gomide for his comments, suggestions, and encouragement.

REFERENCES

[1] S. Brunak and B. Lautrup, *Neural Networks: Computers With Intuition*. Singapore: World Scientific, 1990.
[2] J. Anderson, *An Introduction to Neural Networks*. Cambridge, MA: MIT Press, 1995.
[3] J. Hopfield and D. Tank, "Neural computations of decisions in optimization problems," *Biol. Cybern.*, vol. 52, pp. 141–152, 1985.
[4] M. H. Hassoun, *Associative Neural Memories: Theory and Implemantation*. Oxford, U.K: Oxford Univ. Press, 1993.
[5] M. M. Nelson and W. Illingworth, *A Practical Guide to Neural Nets*. Reading, MA: Addison-Wesley, 1994.
[6] B. Raducanu, M. Grana, and X. F. Albizuri, "Morphological scale spaces and associative morphological memories: Results on robustness and practical applications," *J. Math. Imag. Vision*, vol. 19, no. 2, pp. 113–131, 2003.
[7] B.-L. Zhang, H. Zhang, and S. S. Ge, "Face recognition by applying wavelet subband representation and kernel associative memory," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 166–177, Jan. 2004.
[8] W. Taylor, "Eletrical simulation of some nervous system functional activities," *Inf. Theory*, vol. 3, pp. 314–328, 1956.
[9] K. Steinbuch, "Die lernmatrix ," *Kybernetik*, vol. 1, pp. 36–45, 1961.
[10] J. Anderson, "A simple neural network generating interactive memory," *Math. Biosci.*, vol. 14, pp. 197–220, 1972.
[11] T. Kohonen, "Correlation matrix memory," *IEEE Trans. Comput.*, vol. C-21, no. 4, pp. 353–359, Apr. 1972.
[12] K. Nakano, "Associatron: A model of associative memory," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-2, no. 3, pp. 380–388, Mar. 1972.
[13] T. Kohonen and M. Ruohonen, "Representation of associated data by matrix operators," *IEEE Trans. Comput.*, vol. C-22, no. 7, pp. 701–702, Jul. 1973.
[14] T. Kohonen, *Associative Memory—A System Theoric Approach*. Berlin, Germany: Springer-Verlag, 1977.
[15] ——, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1984.
[16] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. Natl. Acad. Sci.*, Apr. 1982, vol. 79, pp. 2554–2558.
[17] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall, 1999.
[18] B. Kosko, "Adaptive bidirectional associative memories," *Appl. Opt.*, vol. 26, no. 23, pp. 4947–4960, Dec. 1987.
[19] ——, "Bidirectional associative memories," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 49–60, Jan./Feb. 1988.
[20] T. Chiueh and R. Goodman, "Recurrent correlation associative memories," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 275–284, Feb. 1991.
[21] G. X. Ritter, P. Sussner, and J. L. D. de Leon, "Morphological associative memories," *IEEE Trans. Neural Netw.*, vol. 9, no. 2, pp. 281–293, Mar. 1998.
[22] G. X. Ritter, J. L. D. de Leon, and P. Sussner, "Morphological bidirectional associative memories," *Neural Netw.*, vol. 6, no. 12, pp. 851–867, 1999.
[23] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[24] G. X. Ritter and G. Urcid, "Lattice algebra approach to single-neuron computation," *IEEE Trans. Neural Netw.*, vol. 14, no. 2, pp. 282–295, Mar. 2003.
[25] F.-L. Chung and T. Lee, "On fuzzy associative memory with multiple-rule storage capacity," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 4, pp. 375–384, Aug. 1996.
[26] Q. Cheng and Z.-T. Fan, "The stability problem for fuzzy bidirectional associative memories," *Fuzzy Sets Syst.*, vol. 132, no. 1, pp. 83–90, Nov. 2002.
[27] F. Junbo, J. Fan, and S. Yan, "A learning rule for fuzzy associative memories," in *Proc. IEEE Int. Conf. Neural Networks, IEEE World Congr. Computational Intelligence*, Jun.–Jul. 1994, vol. 7, pp. 4273–4277.
[28] P. Liu, "The fuzzy associative memory of max-min fuzzy neural networks with threshold," *Fuzzy Sets Syst.*, vol. 107, pp. 147–157, 1999.
[29] S. T. Wang and H. Lu, "On new fuzzy morphological associative memories," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 3, pp. 316–323, Jun. 2004.
[30] J. P. Choudhury, B. Sarkar, and S. Mukherjee, "Forecasting of engineering manpower through fuzzy associative memory neural network with arima: A comparative study," *Neurocomput.*, vol. 47, pp. 241–257, 2002.
[31] W. Pedrycz, "Neurocomputations in relational systems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 3, pp. 289–296, May 1991.
[32] ——, "Fuzzy neural networks and neurocomputations," *Fuzzy Sets Syst.*, vol. 56, no. 1, pp. 1–28, 1993.
[33] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.
[34] W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design*. Cambridge, MA: MIT Press, 1998.
[35] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, 2nd ed. New York: Chapman & Hall, 1999.
[36] D. Hebb, *The Organization of Behavior*. New York: Wiley, 1949.
[37] J. Drewniak, "Fuzzy relation equations and inequalities," *Fuzzy Sets Syst.*, vol. 14, no. 3, pp. 237–247, Dec. 1984.
[38] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. Venkatesh, "The capacity of the hopfield associative memory," *IEEE Trans. Inf. Theory*, vol. IT-1, no. 1, pp. 33–45, Jan. 1987.
[39] The Database of Faces of AT&T Laboratories Cambridge, U.K. [Online]. Available: http://www.uk.research.att.com/facedatabase.html
[40] G. Birkhoff, *Lattice Theory*, 3rd ed. Providence, RI: AMS, 1993.
[41] P. Sussner, "Fixed points of autoassociative morphological memories," in *Proc. Int. Joint Conf. Neural Networks*, Como, Italy, Jul. 2000, pp. 611–616.
[42] ——, "Associative morphological memories based on variations of the kernel and dual kernel methods," *Neural Netw.*, vol. 16, no. 5, pp. 625–632, Jul. 2003.
[43] R. Cuninghame-Green, *Minimax Algebra: Lecture Notes in Economics and Mathematical Systems 166*. New York: Springer-Verlag, 1979.
[44] ——, "Minimax algebra and applications," in *Advances in Imaging and Electron Physics*. New York: Academic, 1995, vol. 90, pp. 1–121.
[45] G. X. Ritter, J. N. Wilson, and J. L. Davidson, "Image algebra: An overview," *Comput. Vision, Graph., Image Process.*, vol. 49, no. 3, pp. 297–331, Mar. 1990.



Peter Sussner received the Ph.D. degree in mathematics from the University of Florida, Gainesville, partially supported by a Fulbright Scholarship, in 1996.

He is an Assistant Professor in the Department of Applied Mathematics of the State University of Campinas, São Paulo, Brazil. He also acts as a Researcher of the Brazilian National Science Foundation CNPq and holds a membership in the International Neural Network Society. He has previously worked as a researcher at the Center of Computer Vision and Visualization at the University of Florida. He has regularly published articles in refereed international journals, book chapters, and conference proceedings in the areas of artificial neural networks, fuzzy systems, computer vision, mathematical imaging, and global optimization. His current research interests include neural networks, fuzzy systems, mathematical morphology, and lattice algebra.

Marcos Eduardo Valle is working toward the Ph.D. degree in the Department of Applied Mathematics of the State University of Campinas, São Paulo, Brazil.

He holds a scholarship of the Brazilian National Science Foundation CNPq. His current research interests include fuzzy set theory, neural networks, and mathematical morphology.