Minimal Representations and Morphological Associative Memories for Pattern Recall

Gerhard X. Ritter
CISE Department, University of Florida, Gainesville, FL 32611, USA.
Email: ritter@cise.ufl.edu
Gonzalo Urcid S.
Optics Department, INAOE, Tonantzintla, Pue. 72000, Mexico.
Email: gurcid@inaoep.mx

Abstract

Morphological associative memories are a recent development in the field of artificial neural feed-forward networks. The underlying computation of morphological associative memories is based on lattice algebra. Lattice algebra uses the operations of minimum or maximum together with addition, instead of addition and multiplication commonly employed in traditional models of associative memories. In comparison with associative memories based in the Hopfield recurrent network model, morphological associative memories have bigger storage capacity for exemplar patterns and possess better recall capabilities of non-boolean patterns degraded by erosive and dilative noise. Key notions, such as, kernels, morphological independence and minimal representantions are introduced in order to recall stored patterns distorted by random noise.

1 Introduction

The purpose of the present paper is to introduce novel concepts in the theory and applications of *morphological associative memories* (MAM's) as a new mathematical technique for storing image patterns and recalling degraded input versions of the stored patterns. In Section 2, we discuss some concepts and operations of matrix lattice algebra that form the basis for the morphological associative memories. MAM's are defined in a similar way to the classical method of correlation encoding. Section 3 describes the fundamentals of morphological associative memories and their application to the restoration of degraded *non-boolean* patterns by erosive and dilative noise. In Section 4 the key concepts of morphological strong independence and minimal representantions are defined and used for dealing with non-boolean patterns corrupted with random noise. Section 5 gives the conclusions to the present work.

2 Matrix lattice algebra

The basic numerical operations of taking the maximum or minimum of two numbers usually denoted as functions $\max(x,y)$ and $\min(x,y)$ will be written as binary operators using the "join" and "meet" symbols employed in lattice theory [1], i.e., $x \vee y = \max(x,y)$ and $x \wedge y = \min(x,y)$; since both operators are associative, we will use the following symbols to represent the maximum and minimum of a finite set of real numbers $\{x_1, \ldots, x_n\}$

$$\bigvee_{i=1}^{n} x_i = \max_{1\leq i\leq n}\{x_i\} \text{ and } \bigwedge_{i=1}^{n} x_i = \min_{1\leq i\leq n}\{x_i\}. \tag{1}$$

It was shown in [2, 11] that the algebraic systems $(\mathbb{R}_{-\infty}, \vee, +)$ and $(\mathbb{R}_{+\infty}, \wedge, +')$ are *semirings* for corresponding max-$\vee$, min-$\wedge$ operations and their respective additions, $+$ and $+'$ over the sets $\mathbb{R}_{-\infty}$ $\mathbb{R}\cup\{-\infty\}$ and $\mathbb{R}_{\infty} = \mathbb{R}\cup\{+\infty\}$, and that $(\mathbb{R}_{\pm\infty}, \vee, \wedge, +, +')$ is a *bounded lattice-group* or bounded *l*-gro where $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{-\infty, +\infty\}$ is the set of extended real numbers.

Many popular models of associative memories allow for a formulation using matrices [6, 7, 4, 9]. model of associative memories described in this paper can also be expressed using matrix products. $\mathbb{R}$ denote the set of real numbers. For an $m \times p$ matrix A and a $p \times n$ matrix B with entries from $\mathbb{R}$, matrix *product* $C = A \boxed{\vee} B$, also called the *max product* of A and B, is defined by

$$c_{ij} = \bigvee_{k=1}^{p} (a_{ik} + b_{kj}).$$

The *min product* of A and B is defined in a similar fashion. Specifically, the i,jth entry of $C = A \boxed{\wedge} B$ given by

$$c_{ij} = \bigwedge_{k=1}^{p} (a_{ik} + b_{kj}).$$

Note that these matrix products are of the same form as regular matrix products with the role of addition and multiplication replaced by the operations of maximum (or minimum) and addition, respectively. The addition of two matrices is replaced by the operation of maximum (or minimum) in this matrix algebra. Specifically, the maximum of two matrices replaces the usual matrix addition of linear algebra. Here the i,jth entry of the matrix $C = A \vee B$ is given by $c_{ij} = a_{ij} \vee b_{ij}$. Similarly, the minimum two matrices $C = A \wedge B$ is defined by $c_{ij} = a_{ij} \wedge b_{ij}$. The *morphological outer product* of two vectors $\mathbf{x} = (x_1, \cdots, x_n)' \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \cdots, y_m)' \in \mathbb{R}^m$ is defined as

$$\mathbf{y} \times \mathbf{x}' = \begin{pmatrix} y_1 + x_1 & \cdots & y_1 + x_n \\ \vdots & \ddots & \vdots \\ y_m + x_1 & \cdots & y_m + x_n \end{pmatrix}.$$

It is worthwhile to note that $\mathbf{y} \times \mathbf{x}' = \mathbf{y} \boxed{\vee} \mathbf{x}' = \mathbf{y} \boxed{\wedge} \mathbf{x}'$. Finally, we say that A is *less or equal than* denoted by $A \leq B$, and A is *strictly less than* B, denoted by $A < B$, if and only if for each corresponding entry of these matrices we have that $a_{ij} \leq b_{ij}$ and $a_{ij} < b_{ij}$, respectively.

3 Morphological associative memories

3.1 Background

One of the first goals achieved in the development of morphological neural networks was the establishment of a morphological associative memory network (MAM). In its basic form, this model of an associative memory resembles the well-known correlation memory or linear associative memory [6]. As in correlation encoding, the morphological associative memory provides a simple method to add new associations. A weakness in correlation encoding is the requirement of orthogonality of the key vectors in order to exhibit perfect recall of the fundamental associations. The morphological auto-associative memory does not restrict the domain of they key vectors in any way. Thus, as many associations as desired can be encoded into the memory [12].

In the real number case, the capacity for a memory of length n can be as large as desired. That is, if k denotes the number of distinct patterns of length n to be encoded, then k is allowed to be any integer, no matter how large. Of course, in the binary case, the limit is $k = 2^n$ as this is the maximum number of distinct patterns of length n. In comparison, McEliece et al. showed that the asymptotic limit capacity of the Hopfield associative memory is $n/2 \log n$ if with high probability the unique fundamental memory is to be recovered, except for a vanishingly small fraction of fundamental memories [8]. Likewise, the information storage capacity (number of bits which can be stored and recalled associatively) of the morphological auto-associative memory also exceeds the respective number of certain linear matrix associative memories which was calculated by Palm [10]. Among the various auto-associative networks the Hopfield network is the most widely known today [4, 5]. Unlike the Hopfield network, which is a recurrent neural network, the morphological model provides the final result in one pass through the network without any significant amount of training. Ritter, Sussner, and Diaz de Leon used a number of experiments to demonstrate the MAM's efficiency to deal with either dilative (additive) or erosive (subtractive) changes to input (binary) patterns, including incomplete patterns [12]. Thus, the morphological associative memory model provides almost all the characteristics of an ideal associative memory with one notable exception: the network's inability to deal with patterns which contain both erosive as well as dilative noise.

Henceforth, let $(\mathbf{x}^1,\mathbf{y}^1),\ldots,(\mathbf{x}^k,\mathbf{y}^k)$ be k vector pairs with $\mathbf{x}^\xi = (x_1^\xi,\ldots,x_n^\xi)' \in \mathbb{R}^n$ and $\mathbf{y}^\xi = (y_1^\xi,\ldots,y_m^\xi)' \in \mathbb{R}^m$ for $\xi = 1,\ldots,k$. For a given set of pattern associations $\{(\mathbf{x}^\xi,\mathbf{y}^\xi) : \xi = 1,\ldots,k\}$ we define a pair of associated pattern matrices (X,Y), where $X = (\mathbf{x}^1,\ldots,\mathbf{x}^k)$ and $Y = (\mathbf{y}^1,\ldots,\mathbf{y}^k)$. Thus, X is of dimension $n \times k$ with i,jth entry x_i^j and Y is of dimension $m \times k$ with i,jth entry y_i^j.

The earliest neural network approach to associative memories was the linear associative memory or correlation memory [6]. In this approach the goal is to store k vector pairs $(\mathbf{x}^1,\mathbf{y}^1),\ldots,(\mathbf{x}^k,\mathbf{y}^k)$ in an $m \times n$ *associative memory* W such that for any given input vector $\mathbf{x}^\xi$, the associative memory W recalls the output vector $\mathbf{y}^\xi = W\mathbf{x}^\xi, \forall \xi = 1,\ldots,k$. The simplest solution for this goal is to set

$$W = \sum_{\xi=1}^{k} \mathbf{y}^\xi(\mathbf{x}^\xi)'. \tag{5}$$

In this case, the i,jth entry of W is given by $w_{ij} = \sum_{\xi=1}^{k} y_i^\xi x_j^\xi$. If the input patterns $\mathbf{x}^1,\ldots,\mathbf{x}^k$ are orthonormal, then

$$W\mathbf{x}^\xi = \mathbf{y}^\xi((\mathbf{x}^\xi)'\bullet\mathbf{x}^\xi) + \sum_{\gamma\neq\xi} \mathbf{y}^\gamma((\mathbf{x}^\gamma)'\bullet\mathbf{x}^\xi) = \mathbf{y}^\xi. \tag{6}$$

Thus, we have *perfect recall* of the output patterns $\mathbf{y}^1,\ldots,\mathbf{y}^k$. If $\mathbf{x}^1,\ldots,\mathbf{x}^k$ are not orthonormal (as in most realistic cases), then filtering processes using activation functions become necessary in order to retrieve the desired output pattern.

3.2 Definitions and properties

Morphological associative memories are surprisingly similar to these classical correlation memories. With each pair of pattern associations (X,Y) we associate two natural morphological $m \times n$ memories W_{XY} and M_{XY} defined by

$$W_{XY} = \bigwedge_{\xi=1}^{k} [\mathbf{y}^\xi \times (-\mathbf{x}^\xi)'] \ ; \ M_{XY} = \bigvee_{\xi=1}^{k} [\mathbf{y}^\xi \times (-\mathbf{x}^\xi)']. \tag{7}$$

Note the similarities between the definition of the memory given by Eq.(5) and those defined by Eq.(7). Also, a consequence of Eq.(7) is that

$$W_{XY} \boxtimes X \leq Y \leq M_{XY} \boxtimes X. \tag{8}$$

A fundamental relationship between the *canonical* MAM's and other morphological associative memories is given by the next theorem which was proved in [12].

Theorem 3.1. Let (X,Y) denote the associate sets of pattern vector pairs. Whenever there exists pe recall memories A and B such that $A \boxed{\vee} \mathbf{x}^\xi = \mathbf{y}^\xi$ and $B \boxed{\wedge} \mathbf{x}^\xi = \mathbf{y}^\xi$ for $\xi = 1, \ldots, k$, then

$$A \le W_{XY} \le M_{XY} \le B \text{ and}$$
$$\forall \xi, \; W_{XY} \boxed{\vee} \mathbf{x}^\xi = \mathbf{y}^\xi = M_{XY} \boxed{\wedge} \mathbf{x}^\xi.$$
(

Hence, W_{XY} is the least upper bound of all perfect recall memories involving the $\boxed{\vee}$ operation .. M_{XY} is the greatest lower bound of all perfect memories involving the $\boxed{\wedge}$ operation. Furthermore, if the exists perfect recall memories, then the canonical memories are also perfect recall memories. If $X =$ (i.e., $\forall \xi, \mathbf{x}^\xi = \mathbf{y}^\xi$), then we obtain the morphological auto-associative memories W_{XX} and M_{XX}. In [12 we proved that

$$W_{XX} \boxed{\vee} X = X = M_{XX} \boxed{\wedge} X, \quad (10)$$

where X can consist of any arbitrarily large number of pattern vectors. For example, consider the sev

Figure 1: The seven patterns in the top row were used in constructing the memories W_{XX} and M_{XX} (of size 2500×2500). The bottom row shows the output for these morphological autoassociative memories when presented with the respective patterns in the top row.

pattern images $\mathbf{p}^1, \ldots, \mathbf{p}^7$ shown in Fig.(1). Each $\mathbf{p}^\xi$ is a 50×50 pixels 256-gray scale image. For uncorrupted input, perfect recall is guaranteed if we use the memory W_{XX} or M_{XX}. Using the standard row-scan method, each pattern image $\mathbf{p}^\xi$ can be converted into a pattern vector $\mathbf{x}^\xi = (x_1^\xi, \ldots, x_{2500}^\xi)$ by defining

$$x_{50(r-1)+c}^\xi = p^\xi(r, c) \text{ for } r, c = 1, \ldots, 50. \quad (11)$$

3.3 Erosive and dilative noise

Morphological associative memories are extremely robust in the presence of certain types of noise, missing data, or occlusions. We say that a distorted version $\tilde{\mathbf{x}}^\gamma$ of the pattern $\mathbf{x}^\gamma$ has undergone an *erosive change* whenever $\tilde{\mathbf{x}}^\gamma \le \mathbf{x}^\gamma$ and a *dilative change* whenever $\tilde{\mathbf{x}}^\gamma \ge \mathbf{x}^\gamma$. Corrupting the patterns $\mathbf{x}^\xi$ with 30% randomly generated erosive and dilative noise with an intensity level of 128 results in almost perfect recall (NMSE $< 10^{-3}$)[1] when using the memory W_{XX} and M_{XX}, respectively. Figs.(2)–(3) provide for a visual example of this experiment. The reason for the robustness of associative memories in the presence of erosive or dilative noise is a consequence of the following theorem which was validated in [12].

Theorem 3.2. Let $\tilde{\mathbf{x}}^\gamma$ denote the distorted version of the pattern $\mathbf{x}^\gamma$. Then $W_{XY} \boxed{\vee} \tilde{\mathbf{x}}^\gamma = \mathbf{y}^\gamma$ if and only if $\forall j = 1, \ldots, n$

$$\tilde{x}_j^\gamma \le x_j^\gamma \vee \bigwedge_{i=1}^{m} \left(\bigvee_{\xi \ne \gamma} \left[y_i^\gamma - y_i^\xi + x_j^\xi \right] \right) \quad (12)$$

[1] Normalized mean square error, computed for each ξ as $\sum_j (\tilde{x}_j^\xi - x_j^\xi)^2 / \sum_j (x_j^\xi)^2$

Figure 2: Top row: corrupted input patterns (erosive change); bottom row, the corresponding recalled patterns using the morphological memory W_{XX}.

Figure 3: Top row: corrupted input patterns (dilative change); bottom row, the corresponding recalled patterns using the morphological memory M_{XX}.

and for each row index $i \in \{1, \ldots, m\}$ there exists a column index $j_i \in \{1, \ldots, n\}$ such that

$$\tilde{x}^\gamma_{j_i} = x^\gamma_{j_i} \vee \left(\bigvee_{\xi \neq \gamma} \left[y^\gamma_i - y^\xi_i + x^\xi_{j_i} \right] \right). \tag{13}$$

A similar result holds for the memory M_{XY} by changing the inequality sign and the operations of maximum and minimum in Eqs. (12)–(13). The next corollary is an easy consequence of this theorem [14].

Corollary 3.3. Suppose that $\tilde{\mathbf{x}}^\gamma$ denotes an eroded version of $\mathbf{x}^\gamma$. The equation $W_{XX} \boxed{\wedge} \tilde{\mathbf{x}}^\gamma = \mathbf{x}^\gamma$ holds if and only if for each row index $i \in \{1, \ldots, n\}$ there exists a column index $j_i \in \{1, \ldots, n\}$ such that

$$\tilde{x}^\gamma_{j_i} = x^\gamma_{j_i} \vee \left(\bigvee_{\xi \neq \gamma} \left[x^\gamma_i - x^\xi_i + x^\xi_{j_i} \right] \right). \tag{14}$$

A similar result holds for the memory M_{XX} by changing the maximum operator by the minimum operator in Eq. (14).

Although Theorem 3.2 and its Corollary 3.3 provides necessary and sufficient conditions for the bounds of the corruption of the pattern $\mathbf{x}^\gamma$ that guarantees perfect recall, it also implies that W_{XX} will fail if dilative noise not satisfying these bounds is present. Insertion of only minute amounts of dilative noise, often in only one vector component, can result in complete recall failure. Similar comments hold for the memory M_{XX} and erosive noise. Hence, neither memory W_{XX} or M_{XX} is useful in the presence of random noise which, generally, consists of both erosive as well as dilative noise.

The kernel method proposed in [14, 12] suggests a solution to this dilemma. However, it became clear that finding an algorithmic method for selecting an optimal set of proper kernels was not going to be an easy task. Part of the difficulty is due to the fact that the existence of proper kernels for a given set of pattern vectors remains an unsolved problem if the definition of kernels proposed in [12] is used.

More recently, Ritter, Urcid, and Iancu [13] introduced the notion of minimal representations of exemplar *non-boolean* patterns and by redefining the concept of kernels they illustrate that kernels can be effectively used for building morphological associative memories that are robust in the presence of both erosive and dilative noise. In order to deal with the problem of random noise, we introduce, in the next section, the notions of morphological independence and of minimal representations.

4 Morphological independence and minimal representations

4.1 Background

Since W_{XX} is suitable for recognizing patterns corrupted by erosive noise and M_{XX} is suitable for recognizing patterns corrupted by dilative noise, an intuitive idea is to process a noisy version $\tilde{\mathbf{x}}^\gamma$ of $\mathbf{x}^\gamma$ containing both erosive and dilative noise through a combination of W_{XX} and M_{XX}. Sussner proved that passing the output of $M_{XX} \boxdot \tilde{\mathbf{x}}^\gamma$ through the memory W_{XX} or, dually, the output of $W_{XX} \boxdot \tilde{\mathbf{x}}^\gamma$ through M_{XX} will, generally, not result in $\mathbf{x}^\gamma$ [14]. Nevertheless, the modified kernel approach proposed by Ritter et al. is based on this intuitive idea using the memories M_{XX} and W_{XX} in sequence in order to create a morphological memory that is robust in the presence of random noise, even in the general situation where $X \neq Y$ and X and Y are not boolean [12]. The underlying idea is to define a memory M which associates with each input pattern $\mathbf{x}^\gamma$ an intermediate pattern $\mathbf{z}^\gamma$. Another associative memory W is defined which associates each pattern $\mathbf{z}^\gamma$ with the desired output pattern $\mathbf{y}^\gamma$. In terms of min-max products, one obtains the equation:

$$W \boxdot (M \boxdot \mathbf{x}^\gamma) = \mathbf{y}^\gamma. \tag{15}$$

Obviously, if Z is a kernel for (X, Y) and $\mathbf{z}^\gamma \leq \tilde{\mathbf{x}}^\gamma \leq \mathbf{x}^\gamma$, then

$$\mathbf{z}^\gamma = M_{ZZ} \boxdot \mathbf{z}^\gamma \leq M_{ZZ} \boxdot \tilde{\mathbf{x}}^\gamma \leq M_{ZZ} \boxdot \mathbf{x}^\gamma = \mathbf{z}^\gamma \tag{16}$$

and, hence, $M_{ZZ} \boxdot \tilde{\mathbf{x}}^\gamma = \mathbf{z}^\gamma$. Thus, for eroded versions of $\mathbf{x}^\gamma$ that are bounded below by $\mathbf{z}^\gamma \leq \tilde{\mathbf{x}}^\gamma$ we are guaranteed that

$$W_{ZY} \boxdot (M_{ZZ} \boxdot \tilde{\mathbf{x}}^\gamma) = \mathbf{y}^\gamma. \tag{17}$$

Sussner showed the following fundamental result regarding kernels for binary patterns [14].

Theorem 4.1. Let X, Y and Z be sets of binary patterns with $Z \leq X$. If

$$\forall \xi \neq \gamma, \ \mathbf{z}^\gamma \wedge \mathbf{z}^\xi = \mathbf{0} \text{ and } \mathbf{z}^\gamma \not\leq \mathbf{x}^\xi, \tag{18}$$

then Z is a kernel for (X, Y).

4.2 Morphological independence

Various attempts at generalizing this result to the non-boolean case have been unsuccessful. The failure has been due to the fact that in the boolean case the condition specified by Eq.(18) in Theorem 4.1 is implied by the notion of *morphological independence* and results in a kernel. As it turns out, the same is not true in the non-boolean case. A consequence of Sussner's theorem is that for a morphologically independent set of binary patterns X, a kernel set Z for (X, Y) can now be easily chosen. The requirement of morphological independence is not nearly as restrictive as linear independence. Formally, we have:

Definition 4.1. A set of pattern vectors $X = (\mathbf{x}^1, \ldots, \mathbf{x}^k)$ is said to be *morphologically independent* if and only if for $\gamma = 1, \ldots, k$

$$\mathbf{x}^\gamma \not\leq \bigvee_{\xi \neq \gamma} \mathbf{x}^\xi. \tag{19}$$

It is worthwhile to note the resemblance between morphological independence and linear independence. In linear independence, no vector of a set can be a linear sum of the remaining vectors, while in morphological independence, no vector can be less than the maximum of the remaining vectors. This resemblance is due to the algebraic similarity between linear algebra and lattice algebra, where the operation of summation replaced by the operation of maximum and equalities are replaced by appropriate inequalities. We need to mention that morphological independence is a completely different concept than the notion of linear independence in minimax algebra proposed by Cuninghame-Green [3]. A vector can be morphologically independent but not linearly independent as defined by Cuninghame-Green, and vice-versa.

Thus, for example, the set of pattern images shown in Fig.(1) is morphologically independent. It follows from the definition that if X is morphologically independent, then there must be an index $j_\gamma \in \{1,\ldots,n\}$ such that $x^\xi_{j_\gamma} < x^\gamma_{j_\gamma}$ $\forall \xi \neq \gamma$. Hence, $\mathbf{x}^\gamma \not\leq \mathbf{x}^\xi$ $\forall \xi \neq \gamma$. However, the converse does not hold. As mentioned earlier, if X is morphologically independent, then it becomes easy to define a set of patterns $Z \leq X$ that satisfies Eq.(18) of Theorem 4.1. For each $\gamma \in \{1,\ldots,k\}$ we simply pick the index $j_\gamma \in \{1,\ldots,n\}$ for which $x^\xi_{j_\gamma} < x^\gamma_{j_\gamma}$ $\forall \xi \neq \gamma$ and define $\mathbf{z}^\gamma$ by setting

$$z^\gamma_i = \begin{cases} x^\gamma_{j_\gamma} & \text{if } i = j_\gamma \\ 0 & \text{if } i \neq j_\gamma \end{cases} \tag{20}$$

for $i = 1,\ldots,n$. It follows that $\mathbf{z}^\gamma \wedge \mathbf{z}^\xi = \mathbf{0}$ and $\mathbf{z}^\gamma \not\leq \mathbf{x}^\xi$ $\forall\, \xi \neq \gamma$. If X is also boolean, then it follows from Theorem 4.1 that Z is a kernel for (X,Y). However, in our construction of Z we did not assume that X was boolean.

Consider again the pattern images given in Fig.(1), which are morphologically independent. Table 1 gives the row indexes j_γ for $\gamma = 1,\ldots,7$ such that $x^\gamma_{j_\gamma} > x^\xi_{j_\gamma}$ $\forall \xi \neq \gamma$. Using these values of j_γ, the matrix Z was generated using Eq.(20) to test if it resulted in a kernel; Fig.(4) shows in visual form the pixel position corresponding to each pattern row index $j = j_\gamma$ However, Z is not a kernel since the min product

Table 1: Row index j, pixel position (r,c), and value (underscored) x^γ_j of each pattern γ used to build matrix Z.

γ	j	(r,c)	1	2	3	4	5	6	7
1	1759	(36,9)	255	141	165	94	142	159	196
2	453	(10,3)	25	255	71	164	136	101	184
3	2358	(48,8)	233	205	237	163	192	116	107
4	260	(6,10)	20	56	62	255	134	105	44
5	2186	(44,36)	175	112	102	89	255	173	200
6	737	(15,37)	195	70	96	116	136	255	164
7	1276	(26,26)	208	46	199	153	159	176	251

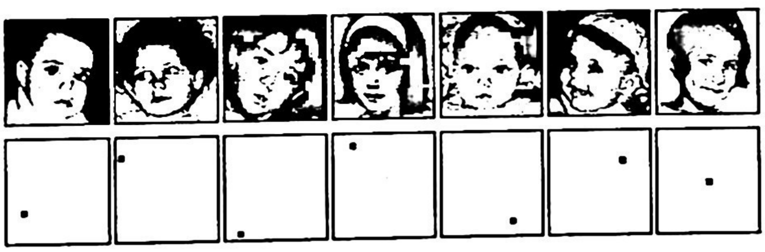

Figure 4: Top row: morphologically independent patterns; bottom row, the corresponding non-zero entries in matrix Z used as a candidate kernel.

$M_{ZZ}\boxed{\wedge}X \neq Z$; also, the max product $W_{XX}\boxed{\vee}(M_{ZZ}\boxed{\wedge}X) \neq X$. Partial reconstruction and the effect of crosstalk noise between patterns for perfect input is shown in Fig.(5); direct computation of $W_{ZX}\boxed{\vee}Z$ is also different from X and visually the output is similar to the bottom row of the same figure.

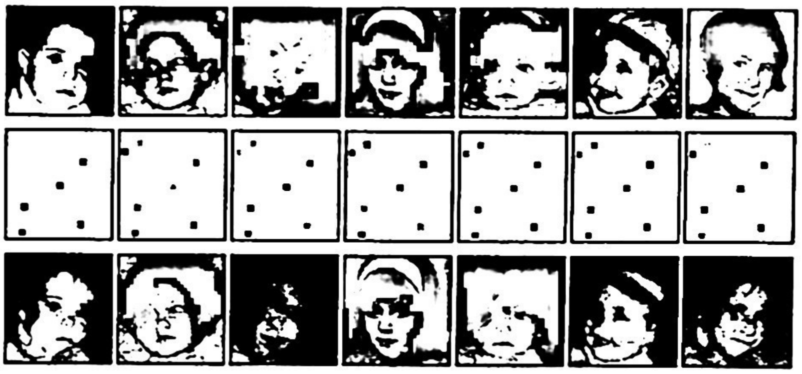

Figure 5: Middle row: output of memory M_{ZZ} when presented with the morphologically independent set shown in the top row; bottom row, output of combined memory scheme $M_{ZZ} \to W_{XX}$.

Definition 4.2. A set of pattern vectors $X = (\mathbf{x}^1, \ldots, \mathbf{x}^k)$ is said to be *morphologically strongly independent* if and only if the following two conditions are satisfied for all $\xi \neq \gamma$:

1. For each $\gamma \in \{1, \ldots, k\}$, $\mathbf{x}^\gamma \not\leq \mathbf{x}^\xi$,
2. For each $\gamma \in \{1, \ldots, k\}$ there exists an index $j_\gamma \in \{1, \ldots, n\}$ such that

$$x^\xi_{j_\gamma} - x^\xi_i \leq x^\gamma_{j_\gamma} - x^\gamma_i, \ \forall i = 1, \ldots, n. \tag{21}$$

The notions of morphological independence and strong independence are, generally not equivalent. However, if X is morphologically independent, then X satisfies Condition 1 of strong morphological independence. Also, we have shown previously, that if X is morphologically independent, then $\forall \gamma = 1, \ldots, k$ there exists an index $j_\gamma \in \{1, \ldots, n\}$ such that $x^\xi_{j_\gamma} < x^\gamma_{j_\gamma}$ $\forall \xi \neq \gamma$. As it turns out, the same property also holds for morphologically strongly independent sets. The proof of the following theorems appears in [13],

Theorem 4.2. Morphological strong independence implies morphological independence.

Theorem 4.3. Suppose X is boolean. Then X is morphologically independent if and only if X is morphologically strongly independent.

Minimal Representations and Kernels

roadblock in obtaining meaningful kernels in the non-boolean case is the overly restrictive requirement that $M_{ZZ}\boxed{\wedge}\mathbf{x}^\gamma = \mathbf{z}^\gamma$. However, if we simply require that there exists a memory W such that

$$W\boxed{\vee}(M_{ZZ}\boxed{\wedge}\mathbf{x}^\gamma) = \mathbf{y}^\gamma, \tag{22}$$

which agrees with our original intuitive idea expressed by Eq.(15), namely that $M_{ZZ}\boxed{\wedge}\mathbf{x}^\gamma$ need only be some intermediate pattern, then several results that mirror Sussner's theorem can be obtained for non-boolean patterns. We therefore suggest the following less restrictive definition of a kernel together with the concept of a minimal representation:

Definition 5.1. Let $Z = (\mathbf{z}^1, \ldots, \mathbf{z}^k)$ be an $n \times k$ matrix. We say that Z is a *kernel* for (X, Y) if and only if $Z \neq X$ and there exists a memory W such that

$$W \boxdot (M_{ZZ} \boxtimes \mathbf{x}^\gamma) = \mathbf{y}^\gamma. \tag{23}$$

If $Y = X$, then we say that Z is a *kernel* for X.

Definition 5.2. A set of patterns $Z \leq X$ is said to be a *minimal representation* of X if and only if for $\gamma = 1, \ldots, k$

1. $\mathbf{z}^\gamma \wedge \mathbf{z}^\xi = \mathbf{0}\ \forall \xi \neq \gamma$,
2. $\mathbf{z}^\gamma$ contains at most one non-zero entry, and
3. $W_{ZX} \boxdot \mathbf{z}^\gamma = \mathbf{x}^\gamma$.

Condition 1 of this definition satisfies part of Eq.(18) of Sussner's theorem while Condition 2 assures sparsity. Condition 3 simply says that X can be reconstructed from Z. In this sense Z acts as an orthogonal basis within the lattice algebra underlying the morphological operations.

There is an obvious close connection between kernels and minimal representations. If Z is a kernel for X in the sense of Definition 4.1, then $Z \leq X$, $M_{ZZ} \boxtimes \mathbf{x}^\gamma = \mathbf{z}^\gamma$, and $W_{ZX} \boxdot \mathbf{z}^\gamma = \mathbf{x}^\gamma$. Thus, kernels satisfy Condition 3 of minimal representations. But from examples given in [12], kernels need not satisfy Conditions 1 and 2 of minimal representations. We now consider the converse, namely for what pattern sets do there exist minimal representations that may also serve as kernels. In the remainder of this section we assume that pattern features are *non-negative*, i.e., $x_i^\gamma \geq 0\ \forall \gamma$ and $\forall i$.

Theorem 5.1. If X is morphologically strongly independent, then there exists a set of patterns $Z \leq X$ with the property that for $\gamma = 1, \ldots, k$

1. $\mathbf{z}^\gamma \wedge \mathbf{z}^\xi = \mathbf{0}\ \forall \xi \neq \gamma$,
2. $\mathbf{z}^\gamma$ contains at most one non-zero entry, and
3. $W_{XX} \boxdot \mathbf{z}^\gamma = \mathbf{x}^\gamma$.

The next corollary is an easy consequence of this theorem.

Corollary 5.2. If X and Z are as in Theorem 5.1, then Z is a minimal representation of X.

Suppose X and Z are as in Theorem 5.1 and $\mathbf{u}^\gamma = M_{ZZ} \boxtimes \mathbf{x}^\gamma$. Then for each $i = 1, \ldots, n$ we have that

$$u_i^\gamma = (M_{ZZ} \boxtimes \mathbf{x}^\gamma)_i = \bigwedge_{j=1}^{n} (m_{ij} + x_j^\gamma) \leq m_{ii} + x_i^\gamma = x_i^\gamma \tag{24}$$

since $m_{ii} = 0$. Hence $\mathbf{u}^\gamma \leq \mathbf{x}^\gamma$ for each $\gamma = 1, \ldots, k$. Since $\mathbf{z}^\gamma \leq \mathbf{x}^\gamma$, it now follows that

$$\mathbf{z}^\gamma = M_{ZZ} \boxtimes \mathbf{z}^\gamma \leq M_{ZZ} \boxtimes \mathbf{x}^\gamma \leq \mathbf{x}^\gamma. \tag{25}$$

In view of Theorem 5.1 and Eq.(25) we have

$$\begin{aligned} \mathbf{x}^\gamma &= W_{XX} \boxdot \mathbf{z}^\gamma = W_{XX} \boxdot (M_{ZZ} \boxtimes \mathbf{z}^\gamma) \\ &\leq W_{XX} \boxdot (M_{ZZ} \boxtimes \mathbf{x}^\gamma) \leq W_{XX} \boxdot \mathbf{x}^\gamma = \mathbf{x}^\gamma. \end{aligned} \tag{26}$$

Therefore,

$$W_{XX} \boxdot (M_{ZZ} \boxtimes \mathbf{x}^\gamma) = \mathbf{x}^\gamma\ \forall \gamma = 1, \ldots, k. \tag{27}$$

By letting $W = W_{XX}$ the preceding argument verifies the following corollary:

Corollary 5.3. If X and Z are as in Theorem 5.1, then Z is a kernel for X.

According to Corollary 5.2, a minimal representation is also a kernel. Hence, for a set of patterns X to be reducible to a kernel, it is sufficient that X is strongly independent. Furthermore, if X is strongly independent, then in order to obtain a kernel one simply selects a minimal representation Z of X using the method given in the proof of Theorem 5.1. Given a minimal representation Z which is also a kernel for X and a noisy version $\tilde{\mathbf{x}}^\gamma$ of the pattern $\mathbf{x}^\gamma$ having the property that $\mathbf{z}^\gamma \leq \tilde{\mathbf{x}}^\gamma$ and $M_{ZZ} \boxtimes \tilde{\mathbf{x}}^\gamma \leq \mathbf{x}^\gamma$, then it must follow that

$$W_{XX} \boxtimes (M_{ZZ} \boxtimes \tilde{\mathbf{x}}^\gamma) = \mathbf{x}^\gamma. \tag{28}$$

The pattern images $\mathbf{p}^1, \ldots, \mathbf{p}^7$ of Fig.(1) were slightly modified so that the new set of patterns is morphologically strongly independent; Table 2 gives the list of the corresponding indexes where the pixel value for each pattern γ in row $j = j_\gamma$ was taken to be the maximum (white) and for $\xi \neq \gamma$ the minimum (black) was assigned. Matrix Z was again defined according to Eq.(18) and applying the results established in Theorem 5.1 and Corollaries 5.2 and 5.3, Z is a minimal representation as well as a kernel respectively; Fig.(6) shows the morphologically strongly independent set of patterns and the associated minimal representation given by $(\mathbf{z}^1, \ldots, \mathbf{z}^7)$. Randomly corrupting the patterns shown in Fig.(6) with 30% of noise

Table 2: Row index j, pixel position (r, c), and value (underscored) x_j^γ of each pattern γ used to build matrix Z.

γ	j	(r,c)	1	2	3	4	5	6	7
1	2463	(50,13)	$\underline{255}$	0	0	0	0	0	0
2	1845	(37,45)	0	$\underline{255}$	0	0	0	0	0
3	2430	(49,30)	0	0	$\underline{255}$	0	0	0	0
4	65	(2,15)	0	0	0	$\underline{255}$	0	0	0
5	112	(3,12)	0	0	0	0	$\underline{255}$	0	0
6	2466	(50,16)	0	0	0	0	0	$\underline{255}$	0
7	14	(1,14)	0	0	0	0	0	0	$\underline{255}$

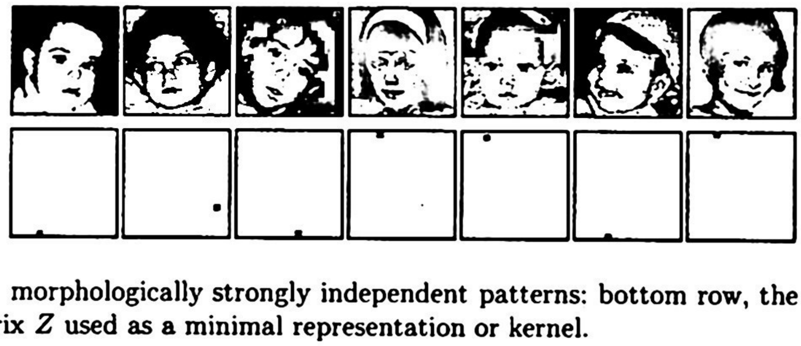

Figure 6: Top row: morphologically strongly independent patterns: bottom row, the corresponding nonzero entries in matrix Z used as a minimal representation or kernel.

with an intensity level of 128, and using the minimal representation Z as our kernel set, we obtained the perfect recall $W_{XX} \boxtimes (M_{ZZ} \boxtimes \tilde{\mathbf{x}}^\gamma) = \mathbf{x}^\gamma$ for $\gamma = 1, \ldots, 7$ shown in Fig.(7).

As we have observed earlier, for a set of patterns X to be reducible to a kernel, it is sufficient that X is strongly independent. Strong independence, however, is not a necessary condition. The question of necessary and sufficient conditions for the existence of kernels remains open. The condition that $W_{XX} \boxtimes \mathbf{z}^\gamma = \mathbf{x}^\gamma$ is crucial in our proof of the kernel scheme

$$input \rightarrow M_{ZZ} \rightarrow W_{XX} \rightarrow output. \tag{29}$$

Figure 7: Top row: corrupted input patterns (random noise); bottom row, perfect recall using the kernel Z for the memory scheme $M_{ZZ} \rightarrow W_{XX}$ (also the output of memory W_{ZX}).

In order to prove Condition 3 of Theorem 5.1, we had to use the fact that X is strongly independent. Thus far we have been unable to weaken the hypothesis of strong independence. There is a good reason why minimal representations are the preferred kernels for the recovery of patterns from noisy inputs. Recall that Eq.(28) will be satisfied whenever

$$z_i^\gamma \leq \tilde{x}_i^\gamma \text{ and } (M_{ZZ} \boxed{\wedge} \tilde{\mathbf{x}}^\gamma)_i \leq x_i^\gamma \tag{30}$$

$\forall i = 1, \ldots, n$. Now, if for some i the i,jth entry of M_{ZZ} is zero for every $j = 1, \ldots, n$, then Eq.(30) is satisfied for this particular index i as long as

$$\bigwedge_{j=1}^{n} \tilde{x}_j^\gamma \leq x_i^\gamma. \tag{31}$$

This claim follows from the fact that since $m_{ij} = \bigvee_{\xi=1}^{k}(z_i^\xi - z_j^\xi) = 0$ for every $j = 1, \ldots, n$ and Z is a minimal representation, we must have that $z_i^\xi = 0$ for every $\xi = 1, \ldots, k$. Hence, $z_i^\gamma = 0 \leq \tilde{x}_i^\gamma$ and

$$(M_{ZZ} \boxed{\wedge} \tilde{\mathbf{x}}^\gamma)_i = \bigwedge_{j=1}^{n}(m_{ij} + \tilde{x}_j^\gamma) = \bigwedge_{j=1}^{n} \tilde{x}_j^\gamma. \tag{32}$$

Due to the lower bound given by Eq.(31), components of $\mathbf{x}^\gamma$ can be arbitrarily corrupted and still satisfy Eq.(30) for the given index i as long as there exists at least one index $j \in \{1, \ldots, n\}$ such that $\tilde{x}_j^\gamma \leq x_i^\gamma$. In many cases, Eq.(30) is automatically satisfied for a large number of indices i whenever Z is a minimal representation. These cases occur when $k \ll n$ as, for example, in the case of the seven image patterns shown in Figs.(1) or (6) where $k = 7 \ll 2500 = n$. If n is large and $k \ll n$, then $n - k$, which is the cardinality of the set $I = \{i : z_i^\xi = 0 \ \forall \xi = 1, \ldots, k\}$, is also large. Since $m_{ij} = \bigvee_{\xi=1}^{k}(z_i^\xi - z_j^\xi)$ and $z_j^\xi > 0$ for at most one ξ, we have that $m_{ij} = 0$ for every $j = 1, \ldots, n$ whenever $i \in I$. This means that M_{ZZ} contains $n - k$ rows having only zero entries. Hence Eq.(30) is satisfied for at least $n - k$ indices i.

Although Eq.(30) is guaranteed to be satisfied for at least $n - k$ indices i, the likelihood that it is satisfied for the remaining k indices is also very high. Since Z is a minimal representation, the inequality $z_i^\gamma \doteq 0 \leq \tilde{x}_i^\gamma$ is guaranteed for all i except one. The only time the inequality may not hold is in the event that for one single index j_γ, $\tilde{x}_{j_\gamma}^\gamma < x_{j_\gamma}^\gamma = z_{j_\gamma}^\gamma$. The probability of this event occurring becomes small as n increases. Also, since M_{ZZ} acts as an erosive memory in that $M_{ZZ} \boxed{\wedge} \tilde{\mathbf{x}}^\gamma \leq \mathbf{x}^\gamma$, the expectation that $(M_{ZZ} \boxed{\wedge} \tilde{\mathbf{x}}^\gamma)_i \leq x_i^\gamma$ is dramatically enhanced for large n.

6 Conclusions

In this paper we describe a new technique for recalling stored patterns from noisy input patterns using morphological associative memories. We define the notions of morphological independence, strong independence, and minimal representations of patterns sets. We also refined the notion of kernels. Our new notion of kernels generalizes the original concept of kernels and provides the major tool for dealing with noisy input pattern when using morphological memories. We established theorems that provide for the existence of minimal representations for strongly independent sets of patterns. We proved that these minimal representations are also kernels and provided the rationale for the preference of these type of kernels over general kernels for the recovery of patterns from noisy inputs. The proof given for the existence of a minimal representation is constructive and provides a method for the construction of kernels. Although we established sufficient conditions for the existence of kernels, we have been unable to establish necessary and sufficient conditions. It is our hope that fellow researchers in this new paradigm of neural computing will be able to solve this problem in the near future.

References

[1] Birkhoff, G. *Lattice Theory.* American Mathematical Society, Providence, RI, 1984.

[2] Cuninghame-Green, R. *Minimax algebra. Lectures Notes in Economics and Mathematical Systems,* Vol. 166. New York, Springer, 1979.

[3] Cuninghame-Green, R. "Minimax algebra and applications," in *Advances in Imaging and Electron Physics,* P. Hawkes, (Ed.), vol. 90, pp.1–121, Academic Press, San Diego, CA, 1995.

[4] Hopfield, J. J. "Neural networks and physical systems with emergent collective computational abilities," in *Proceedings of the National Academy of Sciences,* Vol. 79, pp. 2554–2558, U.S.A., April 1982.

[5] Hopfield, J. J. "Neurons with graded response have collective computational properties like those of two state neurons," in *Proceedings of the National Academy of Sciences,* Vol. 81, pp. 3088–3092, U.S.A., May 1984.

[6] Kohonen, T. "Correlation matrix memory," *IEEE Transactions on Computers,* vol. C-21, pp. 353–359, 1972.

[7] T. Kohonen and M. Ruohonen. "Representation of associative data by computers," *IEEE Transactions on Computers,* vol. C-22, pp. 701–702, 1973.

[8] McEliece, R. et al. "The capacity of Hopfield associative memory," *Trans. Information Theory,* Vol. 1, pp. 33–45, 1987.

[9] Okajima, K., S. Tanaka, and S. Fujiwara. "A heteroassociative memory with feedback connection," in *Proc. IEEE 1st International Conf. on Neural Networks.* Vol. II, pp. 711–718, San Diego, CA, 1987.

[10] Palm, G. "On associative memory," *Biological Cybernetics.* Vol. 36, pp. 19–31, 1980.

[11] Ritter, G. X. *Image Algebra.* Unpublished manuscript available via anonymous ftp from `ftp://ftp.cise.ufl.edu/pub/scr/ia/documents`. University of Florida, Gainesville, FL. 1994.

[12] Ritter, G. X., P. Sussner, and J. L. Diaz de Leon. "Morphological associative memories," *IEEE Transactions on Neural Networks,* Vol. 9(2), pp. 281–293. March 1998.

[13] Ritter, G. X., G. Urcid, and L. Iancu. Reconstruction of noisy patterns using morphological associative memories. Submitted to the *Journal of Mathematical Imaging and Vision.* December 2001.

[14] Sussner, P. "Observations on morphological associative memories and the kernel method." Elsevier *Neurocomputing,* Vol. 31, pp. 167–183. 2000.