A CORRELATION BASED ASSOCIATIVE MEMORY

Dong C. Park and Osama Mohammed
Department of Electrical and Computer Engineering
Florida International University
Miami, FL 33199

S. Oh, Suk Y. Chung, and R. J. Marks II
Department of Electrical Engineering
University of Washington
Seattle, WA 98195

Summary

We propose an associative memory based on minimization of a free energy function determined by the library vectors to be stored. When the library vectors are bipolar, the energy function contains minima at the library vector location. The minima can be sought by search techniques such as gradient descent. Significantly, if the correlation nonlinearity is chosen to be sufficiently strong, then convergence occurs in a single step.

1 Introduction

Associative memory architectures have been presented that are based on artificial neural network structures [1] and iterative matched filters. Indeed, in certain configurations, the resulting procedures produce numerically identical results [2].

Similar associative memories based on nonlinear operations in the correlation domain have also been proposed [3, 4, 5, 6]. Such memories have shift invariant characteristics [4] and high capacity. If the correlation nonlinearity is stronger than $z(x) = N^{x/2}$ and the library vectors are bipolar (*i.e.* contain only ± 1's), then the iterations normally required in such memories need not be used. Convergence, rather, occurs in one step [5, 6].

In this paper, we examine the correlation based associative memory and show its relation to minimization of a free energy function placed on the training data. Using gradient descent, the search is shown to converge in a single step for the case of a bipolar library.

2 Probability of Occurrence and Free Energy

Let $\{\vec{f}_1, \vec{f}_2, \cdots, \vec{f}_N\}$ be library vectors of dimension L that we wish to store in the associative memory. Let $\vec{f}$ be a vector of like dimension. Then, the probability $P(\vec{f})$ is

$$P(\vec{f}) = \sum_{i=1}^{N} P(\vec{f}_i) P(\vec{f} \backslash \vec{f}_i) = \sum_{i=1}^{N} Q_i P(\vec{f} \backslash \vec{f}_i)$$

where $\vec{f}$ is an arbitrary vector and Q_i is the priori probability of the library vector $\vec{f}_i$. We define an energy function $F(\vec{f})$ as

$$F(\vec{f}) = -\gamma \ln[P(\vec{f})] = -\gamma \ln\left[\sum_{i=1}^{N} Q_i P(\vec{f} \backslash \vec{f}_i)\right]$$

where γ is a constant. Note that $F(\cdot)$ is analogous in concept to free energy [8]. Our associative memory paradigm is as follows. Let $\vec{g}$ denote a distorted version of one of the library vectors. With an initialization of $\vec{g}$, we search the landscape of F for a minimum. We will show that the minima occur at library vectors location in many important cases.

EXAMPLE 1 : If $P(\vec{g} \backslash \vec{f}_i)$ has an identically independent Gaussian distribution, then

$$P(\vec{g} \backslash \vec{f}_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\| \vec{g} - \vec{f}_i \|^2}{2\sigma^2}\right].$$

Let $\gamma = 2\sigma^2$ and $a = 1/\gamma$. Then

$$F(\vec{g}) = -\frac{1}{a} \ln\left[\sum_{i=1}^{N} Q_i \exp\{-a \| \vec{g} - \vec{f}_i \|^2\}\right] + C \quad (1)$$

where C is a constant. An example of the energy landscape is illustrated Figure 1 for $a = 8$ and $a = 32$ with $C = 0$. The library vectors are $L = 1$ dimensional and are $\{-1,\ 0,\ 0.5,\ 1\}$ for both values of a. The energy function reaches local minima at the neighborhood of the library vector location when $a = 32$, but this is not the case when $a = 8$. Thus we see that the minima become more distinct for large values of a (which, in the free energy analogy, is inversely proportional to temperature).

EXAMPLE 2 : If all vectors are bipolar (*i.e.* contain only ± 1 components) and the noise is independent flip (Bernoulli) noise with probability $p < 1/2$, then, under the assumption that $\vec{f}$ is bipolar,

$$P(\vec{g} \backslash \vec{f}_i) = p^{k_i}(1-p)^{L-k_i} \quad (2)$$

CH2998-3/91/0000-0901$01.00©1991IEEE

where k_i is the Hamming distance between $\vec{g}$ and $\vec{f}_i$. Equation (2) can be written as

$$P(\vec{g}\backslash\vec{f}_i) = (1-p)^L \left(\frac{p}{1-p}\right)^{k_i} = (1-p)^L \exp[-k_i b] \quad (3)$$

where

$$b = \ln\left(\frac{1-p}{p}\right).$$

and $\| \vec{g} - \vec{f}_i \|^2 = 4k_i$. Equation (3) can then be written as

$$P(\vec{g}\backslash\vec{f}_i) = (1-p)^L \exp[-a \| \vec{g} - \vec{f}_i \|^2]$$

and let $a = b/4$. It follows that the energy function has the same form as (1)

$$F(\vec{g}) = -\gamma \ln\left[\sum_{i=1}^{N} Q_i \exp\{-a \| \vec{g} - \vec{f}_i \|^2\}\right] + C \quad (4)$$

3 Derivation of Operation

Searching for the maximum of $P(\vec{g})$ is equivalent to searching for the minimum of $F(\vec{g})$. The gradient descent method is a popularly used search technique and can be written as

$$\vec{g}(n+1) = \vec{g}(n) - \beta \vec{\nabla}_g F(\vec{g}) \quad (5)$$

where $\vec{\nabla}_g F(\vec{g})$ is gradient of $F(\vec{g})$. If initialization of the search is close to minimum, say $\vec{f}_1$, then we expect that $\vec{g}(\infty) = \vec{f}_1$.

3.1 Continuous Level Library Vector

In Example 1, the gradient is

$$\begin{aligned} \vec{\nabla}_g F(\vec{g}) &= -\frac{1}{a} 2a \frac{\sum_{i=1}^{N} Q_i(\vec{g} - \vec{f}_i)\exp(-a \| \vec{g} - \vec{f}_i \|^2)}{\sum_{i=1}^{N} Q_i \exp(-a \| \vec{g} - \vec{f}_i \|^2)} \\ &= 2\vec{g} - 2\frac{\sum_{i=1}^{N} Q_i \vec{f}_i \exp(-a \| \vec{g} - \vec{f}_i \|^2)}{\sum_{i=1}^{N} Q_i \exp(-a \| \vec{g} - \vec{f}_i \|^2)} \end{aligned}$$

Equation (5) is

$$\begin{aligned} \vec{g}(n+1) &= \vec{g}(n) - \beta \vec{\nabla}_g F(\vec{g}) \\ &= (1-2\beta)\vec{g}(n) + 2\beta \frac{\sum_{i=1}^{N} Q_i \vec{f}_i \exp(-a \| \vec{g} - \vec{f}_i \|^2)}{\sum_{i=1}^{N} Q_i \exp(-a \| \vec{g} - \vec{f}_i \|^2)} \end{aligned}$$

Thus, if $\| \vec{f}_i \|$ is the same for all library vectors and $\beta = 1/2$, then

$$\vec{g}(n+1) = \frac{\sum_{i=1}^{N} Q_i \vec{f}_i \exp[2a\alpha_i(n)]}{\sum_{i=1}^{N} Q_i \exp[2a\alpha_i(n)]}$$

where $\alpha_i(n) = \vec{g}^T \vec{f}_i$ is the correlation of $\vec{g}$ with $\vec{f}_i$.

3.2 Bipolar Library Vector

For the bipolar library with equal priori probability, a sufficiently large value of a and corresponding value of β, assures that the gradient descent method converges in one step, if, after the iteration, we project onto the nearest bipolar vector. For the more general case where the nonlinearity is not sufficiently strong, we calculate from (4).

$$\vec{\nabla}_g F(\vec{g}) = \gamma a \left\{ \vec{g} - \frac{\sum_{i=1}^{N} Q_i \vec{f}_i \exp(2a\alpha_i)}{\sum_{i=1}^{N} Q_i \exp(2a\alpha_i)} \right\}$$

where $\alpha_i = \vec{f}_i^T \vec{g}$. If we choose the gain

$$\beta = \frac{1}{\gamma a}$$

then the recurrent relation will be

$$\vec{g}(n+1) = \frac{\sum_{i=1}^{N} Q_i \vec{f}_i \exp[2a\alpha_i(n)]}{\sum_{i=1}^{N} Q_i \exp[2a\alpha_i(n)]} \quad (6)$$

Generally, $\vec{g}(n+1)$ here is not bipolar, so we project it onto the nearest bipolar vector by the sign($\cdot$) operation

$$\vec{g}(n+1) = \text{sign}\left[\frac{\sum_{i=1}^{N} Q_i \vec{f}_i \exp[2a\alpha_i(n)]}{\sum_{i=1}^{N} Q_i \exp[2a\alpha_i(n)]}\right] \quad (7)$$

The block diagram of this operation is shown in Figure 2, and this operation is the same as the exponential function at the correlation domain in iterative matched filter [5, 6].

3.3 On Step Convergence for Bipolar Library

Consider the case where the priori probability, Q_i, is the same for all library vectors. One step convergence can then be assured for the iteration in (7). Specifically, if $a > \ln(N-1)/2$, the operation in (7) converges to one of the library vector in one step [5, 6]. This condition is the same as requiring that $p < 1/N$. Under this condition, we establish the following

Lemma 1 *If $a > \ln(N-1)/2$, then $F(\vec{g}) > F(\vec{f}_i)$ for $\vec{g} \neq \vec{f}_i$ for all $i = 1, 2, \cdots, N$.*

PROOF :

Let $c_{ik} = \text{h}(\vec{f}_i, \vec{f}_k)$ and $d_k = \text{h}(\vec{g}, \vec{f}_k)$ where $\text{h}(\vec{x}, \vec{y})$ is the Hamming distance between $\vec{x}$ and $\vec{y}$. Let

$$A(\vec{f}_i, \vec{g}) = \sum_{k=1}^{N} \mathrm{e}^{-4ac_{ik}} - \sum_{k=1}^{N} \mathrm{e}^{-4ad_k} \quad (8)$$

Because $\text{h}(\vec{f}_i, \vec{f}_k) \leq \text{h}(\vec{f}_i, \vec{g}) + \text{h}(\vec{f}_k, \vec{g})$,

$$\mathrm{e}^{-4ac_{ik}} \geq \mathrm{e}^{-4ad_i} \cdot \mathrm{e}^{-4ad_k}$$

Equation (8) then becomes

$$\begin{aligned} A(\vec{f}_i, \vec{g}) &\geq 1 + \sum_{k \neq i}^{N} \left[\mathrm{e}^{-4ad_i} \cdot \mathrm{e}^{-4ad_k}\right] - \mathrm{e}^{-4ad_i} \sum_{k \neq i}^{N} \mathrm{e}^{-4ad_k} \\ &= 1 - \sum_{k \neq i}^{N} \mathrm{e}^{-4ad_k} - \mathrm{e}^{-4ad_i}\left[1 - \sum_{k \neq i}^{N} \mathrm{e}^{-4ad_k}\right] \\ &= \left[1 - \mathrm{e}^{-4ad_i}\right]\left[1 - \sum_{k \neq i}^{N} \mathrm{e}^{-4ad_k}\right] \end{aligned}$$

If $a > \ln(N-1)/2$, then, for $n \geq 1$

$$\mathrm{e}^{4an} \leq \frac{1}{N-1}$$

This leads us to conclude that $A(\vec{f}_i, \vec{g}) \geq 0$ and, as a consequence, $F(\vec{g}) > F(\vec{f}_i)$.

Q.E.D.

The above lemma establishes that the free energy has local minima at all library vector locations. One step convergence using (7) attains this local minimum without iteration.

4 Conclusion

In this paper, we found the most probable solution for correlation based associative memories using an energy function equal to the logarithm of the probability. For bipolar library vectors, the associative memory with an exponential nonlinearity in correlation domain minimizes the energy which is proportional to $\ln[P(\vec{f})]$. Also, we showed that one step convergence to the desired local minimum can occur for a bipolar library with equal priori probabilities for sufficiently strong nonlinearity.

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proceedings of the National Academy of Science*, (USA), pp. 2554–2558, 1982.

[2] R. J. Marks II and L. E. Atlas, "Geometrical interpretation of Hopfield's content addressable memory neural network," in *Northcon/88 Conference Record*, (Seattle, Washington), 1988.

[3] G. J. Dunning, E. Marom, Y. Owechko, and B. H. Soffer, "Optical holographic associative memory using a phase conjugate resonator," in *SPIE*, pp. 205–213, 1986.

[4] D. Psaltis and J. Hong, "Shift-invariant optical associative memory," *Optical Engineering*, vol. 26, no. 1, pp. 10–15, 1987.

[5] R. J. Marks II, L. E. Atlas, J. J. Choi, S. Oh, K. F. Cheung, and D. C. Park, "Performance analysis of associative memories with nonlinearity in the correlation domain," *Applied Optics*, vol. 27, no. 14, pp. 2900–2904, 1988.

[6] R. J. Marks II, L. E. Atlas, J. J. Choi, S. Oh, K. F. Cheung, and D. C. Park, "Nonlinearity requirements for correlation based associative," in *SPIE*, pp. 179–183, 1988.

[7] T. Chiueh and R. M. Goodman, "High-capacity exponential associative memories," in *Proceedings of the IEEE International Conference on Neural Networks*, (San Diego, California), pp. I 153–I 159, 1988.

[8] F. Reif, *Fundamentals of Statistical and Thermal Physics*. New York: McGraw-Hill, 1965.

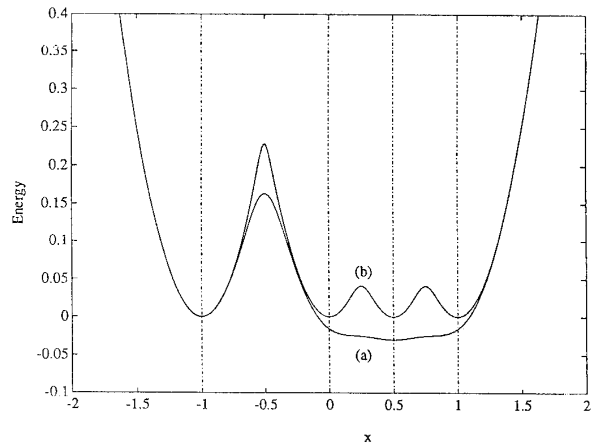

Figure 1 : The energy in the Gaussian noise for (a) $a = 8$ and (b) $a = 32$. The library vectors (scalars in this example) are -1, 0, 0.5, and 1.

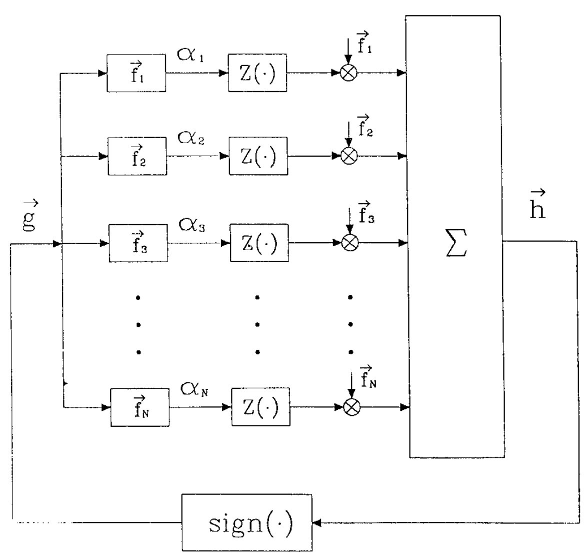

Figure 2 : Block diagram of a matched filter with a nonlinearity in the correlation domain. $Z(x) = e^{2ax}$.

Dong Chul Park (S'80-M'90) received his B.S. degree in 1980 from Sogang University, M.S. degree in 1982 from Korea Advanced Institute of Science and Technology, Seoul, Korea, and Ph.D. degree at the University of Washington, Seattle, in 1990. He joined the faculty of the Department of Electrical and Computer Engineering at the Florida International University, Miami, in August of 1990. He is a committee member for *The First International Forum on Application of Neural Networks to Power Systems.* Since 1988, he has published more than 20 papers including 10 archival journal papers in the areas of artificial neural network, adaptive algorithm for neural networks, electric load forecasting using neural networks, signal processing and optical computing.

Osama Mohammed (S'80, M'83, SM'87) received the MS and Ph.D. degrees in Electrical Engineering in 1980 and 1983 respectively from VPI and SU. Dr. Mohammed has been Assistant, Associate Professor and is currently a Professor in the Department of Electrical and COmputer Engineering, Florida International University, Miami, Florida. Dr. Mohammed's current research interests include techniques for modeling and of electrical power system utilizing artificial intelligence, electromagnetic field modeling in electrical and electronic devices, as well as biomedical applications. He is the author of more than 90 IEEE Transaction, Journals and proceedings. Dr. Mohammed is a member of several IEEE/PES committees. He is also the organizer and director of short courses offered, at FIU, for power system engineers and supervising groups from electric utilities. Dr. Mohammed currently serves as the Chairman of the Miami section of IEEE and is a member of the Florida Power and Light Consumer Advisory Board.