PUBLISHED IN THE
PROCEEDINGS OF THE THIRD IEEE INTERNATIONAL CONFERENCE
ON ELECTRONICS, CIRCUITS, AND SYSTEMS (ICECS'96)
RODOS, GREECE, 13–16 OCTOBER, 1996.

A METHOD FOR ALL-POSITIVE OPTICAL MULTILAYER PERCEPTRONS

I. Saxena *E. Fiesler* *P. Moerland*
IDIAP, CP 592, CH-1920 Martigny, Switzerland, {isaxena,efiesler,perry}@idiap.ch

ABSTRACT

The most promising approaches for optical neural networks are based on intensity encoding. However, a serious drawback of intensity encoding is the lack of negative values and optical subtraction, which are essential for rendering neural networks useful. To overcome the need for optical subtraction, a novel training method is described here that is especially useful for optical multilayer neural networks. In this method, subtraction is implemented as a transformation of the interconnection weights which makes possible the implementation of multilayer perceptrons with optical thresholding. The method is straightforward to implement in optical neural networks where learning occurs under external electronic/computer control.

1. INTRODUCTION

An important feature of neural network models is their massive parallelism of weighted interconnections which connect one or more layers of non-linear processing elements (*neurons*). The basic operation in most neural networks is the calculation of a weighted sum, the outcome of which is transformed by a non-linear thresholding function. The popular multilayer perceptron, for example, consists of several layers performing these two operations in a feedforward fashion from the layer of inputs to the layer of outputs. A sketch of this topology and the associated terminology is given in Figure 1. It is well-known that these networks can approximate arbitrarily well any continuous functional mapping and can learn the weight values suitable for a given data set; see for example [1]. The virtually unlimited range of applications of neural networks and their massive parallelism has fostered the development of optoelectronic and optical neural network implementations. Most optical implementations are incoherent systems based on intensity encoding which are lacking optical subtraction. A straightforward implementation of an intensity based all-optical system would therefore lead to *all-positive* neural networks. These are extremely limited in their performance; for example, a multilayer perceptron that is restricted to the all-positive domain and uses the standard monotonic non-linear thresholding function, can only increase values and never decrease them. This renders the functionality of these all-positive multilayer perceptrons too limited to be useful. Hence, being able to avail of either negative values or subtraction is essential wto the optical implementation of neural networks.

Optical subtraction has only been demonstrated in relatively simple optical systems, but not yet

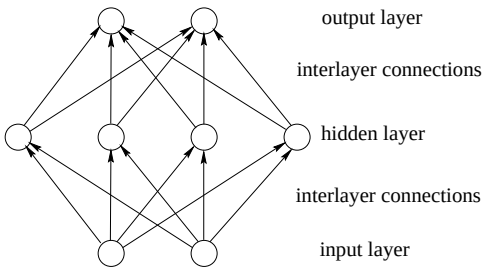

Figure 1. Multilayer Perceptron.

in massively parallel optical architectures, like optical neural networks. Moreover, most optical subtraction methods are based on interferometric subtraction which is difficult to implement, as phase changes are very sensitive to mechanical disturbances. Alternatives to the incoherent intensity encoding approach, using other encoding schemes, like those based on phase or wavelength encoding, are virtually virgin terrain.

Given these limitations, subtraction is usually realized as an electronic difference of two photodetected quantities whose separation is based either on two states of a characteristic parameter of light (polarization [2]) or on spatial or temporal separation [3]. Such photo-electric conversion and subsequent regeneration at optical sources, is not an optimal solution as the light propagation gets interrupted, preventing all-optical neural processing at hidden layers.

The use of both monotonically increasing and decreasing activation functions in the same network [4][5] has been proposed as a way to overcome the subtraction problem. However, the subtraction of biases is still required in this method. Another proposal is the variable thresholding approach [6] for Hopfield nets, which in principle overcomes the need for subtraction. However, to be more generally applicable, its realization would require optical devices with suitably matched characteristics. Recent work in this area [7] has also been motivated by the encumbrance of performing subtraction electronically in optical neural networks (ONNs). Their implementation still requires subtraction of problem-dependent biases and does not permit optical thresholding.

The ideal solution would be to incorporate subtraction in incoherent ONNs in a practical way. Section 2 describes such a solution, which is based on a mathematical technique to avoid optical sub-

© IEEE 1996

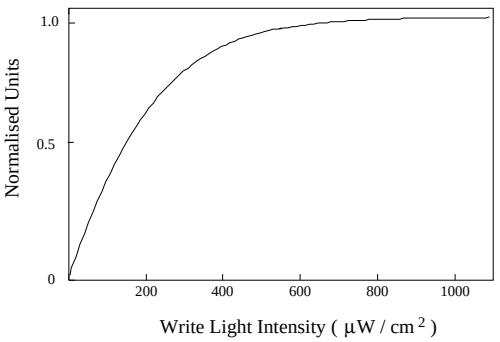

Figure 2. Non-linear response of a liquid crystal light valve.

traction and provides a scheme for implementing all-positive neural networks with optical thresholding that are trained under electronic/computer control. The described solution is based on a transformation of the bipolar weights and biases to the positive domain, which can be performed on a computer or even partly on an optical system. This solution is especially consistent with the use of all-positive sigmoid-like activation functions, and hence enables multilayeredness in optical neural networks. A range of experiments on training neural networks evaluates the viability of the proposed solution.

2. SUBTRACTION COMPENSATION

In this section a method to obtain an all-positive ONN is described. In an optical implementation of neural networks, for which a computer is used for the training phase, this method enables thresholding by an optical non-linear response [8][9], permitting all-optical forward propagation in the ONN.

2.1. Theory

In the forward propagation pass of a typical neural network model, each hidden and output neuron assumes an activation value a_j given by:

$$a_j = S\left(\sum_i W_{ij}a_i - \theta_j\right).$$

The hidden layer neuron firstly obtains the difference of the weighted sum, $\sum_i W_{ij}a_i$, with the bias, θ_j, followed by a non-linear thresholding $S()$ of this difference. In the method described, the thresholding, $S()$, is done optically, where $S(x)$ is defined for $x \geq 0$ [8]. Figure 2 shows an example of such an all-positive non-linear response curve of a liquid crystal light valve.

To provide an all-positive argument to the non-linear function $S()$, it is desired to obtain the all-positive equivalent of the difference $(\sum_i W_{ij}a_i) - \theta_j$ where both W_{ij} and θ_j are unconstrained or bipolar, and the a_i are positive. Transformed all-positive weights W'_{ij} are obtained as:

$$W'_{ij} = W_{ij} - W_{min},$$

where W_{min} is the minimum over all weights W_{ij} and bias values θ_j. For any value of W_{min} all the W'_{ij} are, therefore, always rendered positive. The transformation to all-positive weights then results in:

$$\begin{aligned}\sum_i W_{ij}a_i - \theta_j &= \sum_i (W'_{ij} + W_{min})a_i - \theta_j \\ &= \sum_i W'_{ij}a_i - \theta'_j,\end{aligned}$$

where

$$\theta'_j = \theta_j - W_{min}\sum_i a_i. \tag{1}$$

Although the weights are now all-positive, this transformation nevertheless still requires the difference of the new sum (of all-positive weighted inputs) and the new biases. Now, consider:

$$\begin{aligned}\sum_i W_{ij}a_i - \theta_j &= \sum_i W'_{ij}a_i \left(1 - \frac{\theta'_j}{\sum_i W'_{ij}a_i}\right) \\ &= \sum_i W'_{ij}a_i(1 - r_j),\end{aligned}$$

where

$$r_j = \frac{\theta'_j}{\sum_i W'_{ij}a_i}. \tag{2}$$

Note that r_j is $r_j(a_i)$, a function of the a_i. For $0 \leq (1 - r_j)$, the new weights W''_{ij} are all-positive:

$$W''_{ij} = W'_{ij}(1 - r_j),$$

and the all-positive argument of the non-linear function $S()$ is

$$\sum_i W_{ij}a_i - \theta_j = \sum_i W''_{ij}a_i.$$

The effect of the requirement: $0 \leq (1 - r_j)$, is discussed in the next section, while implementation issues are discussed in section 3.

2.2. Discussion

The condition $(1 - r_j) \geq 0$ needs to be fulfilled for $W''_{ij} \geq 0$ to allow an all-positive optical implementation. It is, therefore, required that

$$r_j \leq 1, \tag{3}$$

that is,

$$\frac{\theta'_j}{\sum_i W'_{ij}a_i} \leq 1$$

or

$$\theta'_j \leq \sum_i W'_{ij}a_i.$$

From equation (1),

$$\theta_j \leq \sum_i a_i(W_{min} + W'_{ij})$$

or

$$\theta_j \leq \sum_i W_{ij}a_i.$$

This can be re-written as:

$$\sum_i W_{ij}a_i - \theta_j \geq 0.$$

This condition is consistent with the optical thresholding function definition, $S()$, which is

defined only for all-positive arguments, and for which arguments less than zero are truncated to zero. Hence, the implicit truncation by the optical non-linear function, as in figure 1, is equivalent to the condition of equation (3) being satisfied, that is, a truncation of W''_{ij} to zero.

benchmark	network topology[1]	pattern set sizes		
		train.	valid.	test
XOR	2-4-1	4	-	-
Sonar	60-8-2	104	-	-
Wine	13-6-3	89	44	45
Cancer	9-6-2	350	175	174

Table 1. Summary of the benchmarks used in the experiments.

	#Runs	ε	learning rate	initial weights
XOR	50	0.1	0.3	$[-1, 1]$
Sonar	10	0.3	0.1	$[-1, 1]$
Wine	10	-	0.3	$[-0.5, 0.5]$
Cancer	10	-	0.3	$[-0.5, 0.5]$

Table 2. Summary of the parameters used in the experiments.

To investigate the actual influence on neural network training of a non-linear thresholding function, that is truncated to be all-positive, a series of experiments have been performed. The non-linear thresholding function used in the simulations is a sigmoid that has been translated by 2.75 along the positive x-axis. This sigmoid has a y-intercept of 0.06, which corresponds to the maximum y-intercept of the transfer curves of 4 different liquid crystal light valves available [8]. The experiments consist of a backpropagation[2] (BP) training of a three-layer[3] neural network, using this translated sigmoid with and without truncation at the origin. A set of benchmarks including three real-world problems and the eXclusive OR (XOR) problem has been used. The benchmark characteristics and the training parameters used are listed in Table 1 and 2 respectively. Two benchmarks (XOR and Sonar) have been used to evaluate a straightforward backpropagation training, while the other two benchmarks (Wine and Cancer) have been used to assess generalization performance on a test set, which is not used during training. The criterion for training convergence was based on a maximal difference of ε between all the target and actual network outputs for all training patterns. For the benchmarks that involve generalization, a cross-validation technique was used to decide on when to stop training. This technique involves a training, validation, and test set. A more detailed description of the benchmark problems and the simulation conditions can be found in [12].

The simulation results are outlined in Table 3 for the untruncated sigmoid and in Table 4 for the truncated sigmoid. All the results are averaged over a certain number of runs (#Runs in Table 2) with different random weight initializations. As was to be expected there are some differences between the results for the untruncated and truncated sigmoid.

[1] # of neurons in the input-hidden-output layer. Figure 1 shows, for example, a 2-4-2 topology.

[2] See, for example, [10].

[3] A layer is defined to be a layer of neurons [11].

	#Iterations	%Conv.	%Miscl.
XOR	1403.9	100.0	-
Sonar	859.8	80.0	-
Wine	816.5	-	2.95
Cancer	41.0	-	1.32

Table 3. Results without truncation to zero. #Iterations: number of BP training iterations. %Conv.: percentage of converged runs. %Miscl.: percentage of misclassified test patterns.

	#Iterations	%Conv.	%Miscl.
XOR	1404.6	100.0	-
Sonar	791.4	90.0	-
Wine	125.5	-	2.73
Cancer	40.0	-	1.72

Table 4. Results with a truncation to zero. #Iterations: number of BP training iterations. %Conv.: percentage of converged runs. %Miscl.: percentage of misclassified test patterns.

However, they illustrate that the use of a non-linear thresholding function that is limited to the positive domain, does not deteriorate neural network training.

3. IMPLEMENTATION OF SUBTRACTION COMPENSATION

An adaptive multilayer optical neural network which enables all-optical forward propagation and learning under the control of a computer has been described by Saxena and Fiesler [8]. The optical system uses liquid crystal televisions (LCTVs) to implement the matrix-vector multiplication of a weight matrix and an input vector a_i, whereas liquid crystal light valves (LCLVs) are used to implement non-linear thresholding. See figure 2 for a typical LCLV response curve. This optical system is based on intensity encoding of light and therefore restricted to all-positive computation. It represents a system in which the transformation described in section 2 can be applied to compensate for the lack of subtraction and negative values.

The all-positive weights W''_{ij} can easily be obtained by carrying out the transformations on a host computer and can then be implemented on a weight implementation device, like an LCTV. A part of the transformation, namely the intermediary matrix-vector product $\sum_i W'_{ij} a_i$, needed for the computation r_j by equation (2), may also be obtained optically. A micro-lenslet array can be used to collect the light intensity corresponding to elements of this optical matrix-vector product. Using this result for $\sum_i W'_{ij} a_i$, the term $(1 - r_j)$ and the all-positive weights W''_{ij} can be calculated on a host computer. When the W''_{ij} are obtained, they replace, with a proper rescaling based on the maximal value of W''_{ij}, the W'_{ij} in the weight implementation device, such as an LCTV.

The advantage of the optical over the computational method of calculating $\sum_i W'_{ij} a_i$ would depend on experimental conditions, such as the number of neurons in the network and the temporal responses of the devices used.

4. CONCLUSIONS

A novel method for subtraction compensation is described that enables all-positive forward propaga-

tion in optical neural networks. The resulting all-positive multilayer optical neural networks can be implemented with optical non-linear thresholding, having electronic or computer controlled learning. It allows uninterrupted forward propagation of light in multilayer optical neural networks. Successful neural network training performance is obtained in experiments on four benchmark problems with truncation of the transformed weights to zero.

ACKNOWLEDGEMENTS

The authors would like to thank the Swiss National Science Foundation (FN:21-36497.92) for their partial support of this research.

REFERENCES

[1] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, 1995.

[2] M. Kranzdorf, B. J. Bigner, L. Zhang and K. M. Johnson,"Optical Connectionist Machine with Polarization-Based Bipolar Weight Values," *Optical Engineering*, vol. 28, no. 8, pp. 844-848, Society of Photo-Optical Instrumentation Engineers, August 1989.

[3] N. Farhat and D. Psaltis, "New Approach to Optical Information Processing Based on Hopfield Model," *Journal of the Optical Society of America*, vol. A1, p. 1337, 1984.

[4] F. M. Dickey and J. M. DeLaurentis, "Optical Neural Networks with Unipolar Weights," *Optics Communications*, vol. 101, no. 5/6, pp. 303–305, North-Holland/Elsevier, 1993.

[5] J. M. DeLaurentis and F. M. Dickey,"A Convexity Based Analysis of Neural Networks," *Neural Networks*, vol. 7, no. 1, pp. 141–146, Pergamon Press, 1994.

[6] H. J. White and W. A. Wright, "Holographic Implementations of a Hopfield Model with Discrete Weights," *Applied Optics*, vol. 27, no. 2, pp. 331–338, January 1988.

[7] Y. Hayasaki, I. Tohyama, M. Mori, and S. Ishihara, "Reversal-Input Superposing Technique for All-Optical Neural Networks," *Applied Optics*, vol. 33, no. 8, pp. 1477–1484, 1994.

[8] I. Saxena and E. Fiesler, "Adaptive Multilayer Optical Neural Network with Optical Thresholding," *Optical Engineering*, vol. 34, no. 8, pp. 2435–2440, August 1995.

[9] P. Moerland, E. Fiesler, and I. Saxena, "The Effects of Optical Thresholding in Backpropagation Neural Networks," *Proceedings of the International Conference on Artificial Neural Networks (ICANN'95)*, vol. 2, pp. 339–343, 1995.

[10] D. Rumelhart, G. Hinton, and R. Williams, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Massachusetts: MIT Press, vol. 1: Foundations, pp. 318–362, 1986.

[11] E. Fiesler. "Neural Network Topologies," in *The Handbook of Neural Computation*, E. Fiesler and R. Beale (Editors-in-Chief), Oxford University Press and IOP Publishing, 1996. (ISBN: 0-19-509138-8).

[12] P. Moerland, E. Fiesler, and I. Saxena, "Discrete Multilayer Perceptrons for All-Optical Implementation," in preparation for submission to *Optical Engineering*, 1996.