The Multivalued and Continuous Perceptrons

George M. Georgiou
Computer Science Department
California State University
San Bernardino, CA 92407
E-mail: `georgiou@wiley.csusb.edu`
January, 1993

Abstract

Rosenblatt's perceptron is extended to (1) a multivalued perceptron and (2) to a continuous-valued perceptron. It shown that any function that can be represented by the multivalued perceptron can be learned in a finite number of steps, and any function that can represented by the continuous perceptron can be learned with arbitrary accuracy in a finite number of steps. The whole apparatus is defined in the complex domain. With these perceptrons learnability is extended to more complicated functions than the usual linearly separable ones. The complex domain promises to be a fertile ground for neural networks research.

1 Rosenblatt's Perceptron

One of the most remarkable and important theorems in the neural networks literature is Rosenblatt's Perceptron Theorem [Ros62]. Simply stated, it says that if there is a weight vector $\mathbf{W}^*$ such that $\mathbf{W}^{*T}\mathbf{X} > 0$ whenever pattern vector $\mathbf{X}$ belongs to class ω_1 and $\mathbf{W}^{*T}\mathbf{X} < 0$ whenever X belongs to class ω_0, i.e. the two classes are linearly separable, then a given simple learning algorithm will find a weight vector $\mathbf{W}$ that will satisfy the above two inequalities. The superscript T indicates matrix transpose. In other words, if the pattern vectors, which belong to two classes ω_0 and ω_1, are linearly separable, then the algorithm will find a hyperplane that separates the two classes. The activation function used is the hard thresholding function, which gives two outputs: -1 or 1.

In this work Rosenblatt's Perceptron Theorem is generalized to (1) the Multivalued Perceptron, where the output can take more than two discrete values, and (2) the Continuous Perceptron, where the output takes continuous values. In both cases all quantities involved are complex-valued. The output of the activation functions falls on the unit circle in the complex plane. Rosenblatt's perceptron is a special case of both perceptrons. The activation functions were previously used in [Noe88a] and [Noe88b] in a different context, and are different from the one used for the complex backpropagation algorithm in [Geo92b, GK92].

2 The pattern and weight space

Each input pattern vector is represented by $\mathbf{X} = (x_1, x_2, \ldots, x_n, 1)^T$, where each component $x_i, 1 \le i \le n$, is a complex number, i.e. $x_i \in \mathbb{C}$. The extra component having value 1 is to allow treatment of the threshold as just another weight – an established practice. Thus, $\mathbf{X} \in \mathbb{C}^{n+1}$, but the actual dimension of the input patterns is n. The weight vector is denoted by $\mathbf{W} = (w_1, w_2, \ldots, w_{n+1})$, $w_i \in \mathbb{C}, 1 \le i \le n+1$. Weight w_{n+1} is the (complex) threshold. The complex conjugate of a vector $\mathbf{C} = (c_1, c_2, \ldots, c_{n+1}) \in \mathbb{C}^{n+1}$ is defined as $\overline{\mathbf{C}} = (\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_{n+1})$. The absolute value or magnitude of a complex number $z = u + i\,v$, is $|z| = \sqrt{u^2+v^2}$, where $u = \mathrm{Re}(z)$ and $v = \mathrm{Im}(z)$ are the real and imaginary parts of z respectively.

The scalar product $(\mathbf{W}|\mathbf{X})$ of vectors $\mathbf{W}, \mathbf{X} \in \mathbb{C}^{n+1}$ is defined as the complex number $(\mathbf{W}|\mathbf{X}) = \mathbf{W}^T\overline{\mathbf{X}}$. The norm, or length, of a vector $\mathbf{W} \in \mathbb{C}^{n+1}$ is defined as the non-negative real number $||\mathbf{W}|| = \sqrt{(\mathbf{W}|\mathbf{W})}$. Thus, $||\mathbf{W}||^2 = \sum_{j=1}^{n+1} w_j\overline{w}_j$. An immediate result of this definition is the fact that the norm of two complex conjugate vectors is equal: $||\mathbf{W}|| = ||\overline{\mathbf{W}}||$.

The above definitions are the standard ones for the finite-dimensional complex Hilbert space $\mathbb{C}^{n+1}$ [Ber61].

3 The Multivalued Perceptron

The weights of the Multivalued (q-state) Perceptron are complex-valued and so are its inputs. The output s of the $q-$state perceptron, $q \geq 2$, takes values from the set $\{\sigma^k \mid \sigma^k = e^{i\frac{2\pi k}{q}}$ and $k = 0, 1, 2, \ldots, q-1\}$. Put in other words, the output belongs to the set of the q roots of unity, which are evenly distributed over the unit circle on the complex plane. Naturally, 1 will always be in the set. Each output σ^k corresponds to class ω_k. Figure 1 shows the possible states of the 8−state perceptron, which are shown as arrows.

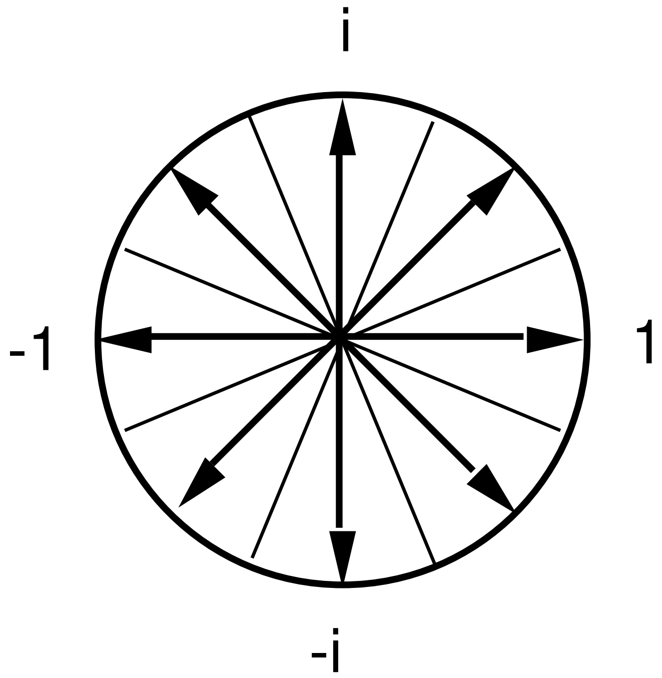

Figure 1: The outputs of the 8−state perceptron are the 8 roots of unity.

Let h_j indicate the net input to the perceptron for input pattern vector $\mathbf{X}$:

$$h_j = \mathbf{W}^T \mathbf{X}_j \tag{1}$$

The output s_j is determined as follows:

$$s_j = \sigma^k \text{ if } \left| \arg(\frac{h_j}{\sigma^k}) \right| < \frac{\pi}{q} \tag{2}$$

Where $\arg(x)$, the phase angle of x, is taken to be in the interval $(-\pi, \pi)$. Thus, the output takes the value of the $q-$th root of unity that is closer (in phase angle) to the net input. Boundary conditions occur when the following condition holds:

$$\left| \arg(\frac{h_j}{\sigma^k}) \right| = \frac{\pi}{q} \tag{3}$$

and can be handled in a consistent manner [Geo92b]. The error ε_j for $\mathbf{X}_j$ is defined as

$$\varepsilon_j \quad = \quad \text{(desired output) - (actual output)} \tag{4}$$

To facilitate the proofs, it is convenient to cast the problem in a slightly different form. If input vector $\mathbf{X} \in \omega_k$ (which means that that the desired output is σ^k), then transform it to $\mathbf{X}' = e^{-i\frac{2\pi k}{q}}\mathbf{X}$. Note that the multiplication is component-wise. This in effect rotates the desired output to $1 = \sigma^0$. After we find $\mathbf{W}$ such that the output every transformed input vector $\mathbf{X}'$ is 1, we can restore each transformed input vector to its original value: $\mathbf{X} = e^{i\frac{2\pi k}{q}}\mathbf{X}'$. This will rotate the output 1 to the desired output σ^k.

From now on we assume that the input vectors are appropriately transformed as explained above. Thus, we are searching for weight vector $\mathbf{W}$ that would yield output 1 for every (transformed) input vector.

Since the desired output is 1 for all vectors, then the error $\varepsilon_j = 1 - s_j$. If we present the input vector $\mathbf{X}_j$ to the q-state Perceptron and the output $s_j \neq 1$, then vector $\mathbf{X}_j$ is not correctly classified. If $s_j = 1$, then $\mathbf{X}_j$ is correctly classified.

We now state the Multivalued (q-state) Perceptron Theorem.

Theorem 1 Starting from a random initialization of the weight vector $\mathbf{W}(1)$, each input pattern $\mathbf{X}_i$ is repeatedly presented at finite time intervals to the q-state Perceptron. At step k of the algorithm, the weight vector is modified according to the rule:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \alpha\, \varepsilon_k \overline{\mathbf{X}}(k) \tag{5}$$

where α is a real positive constant and ε_k the error associated with $\mathbf{X}(k)$. If the vector is correctly classified, obviously there is no change in the weight vector. If there exists $\mathbf{W}^*$ such that all input patterns are correctly classified, then the algorithm will converge in a finite number of steps; that is it will find a weight vector that will correctly classify all input pattern vectors. This will cause the weight vector to stop being modified.

To prove the theorem, we need two lemmas.

Lemma 1 If a weight vector $\mathbf{W}$ correctly classifies a vector $\mathbf{X}_j$ and $\varepsilon \neq 0$ then

$$\mathrm{Re}(\overline{\varepsilon}\mathbf{W}^T\mathbf{X}_j) > 0 \tag{6}$$

Proof: Suppose that $\mathbf{W}$ correctly classifies $\mathbf{X}$. Let $\mathbf{W}^T\mathbf{X}_j = a(\cos(\theta) + i\, \sin(\theta))$ with $a > 0$ and $|\theta| < \frac{\pi}{q}$.

$$\begin{aligned}
\mathrm{Re}(\overline{\varepsilon}\mathbf{W}^T\mathbf{X}_j) &= \mathrm{Re}(\varepsilon)a\cos(\theta) + \mathrm{Im}(\varepsilon)a\sin(\theta) \\
&= a((1-\cos(\frac{2\pi l}{q}))\cos(\theta) - \sin(\frac{2\pi l}{q})\sin(\theta)) \\
&= a(\cos(\theta) - \cos(\theta)\cos(\frac{2\pi l}{q}) - \sin(\theta)\sin(\frac{2\pi l}{q})) \\
&= a(\cos(\theta) - \cos(\frac{2\pi l}{q} - \theta)) \\
&= 2a\sin(\frac{\pi l}{q})\sin(\frac{\pi l}{q} - \theta) \qquad (7)
\end{aligned}$$

All we have to show is that the arguments of the two sine functions are on the open interval $(0, \pi)$, and thus always positive. Since $l = 1, 2, \ldots, q-1$, from $\varepsilon \neq 0$, then

$$\frac{\pi}{q} \leq \frac{\pi l}{q} \leq \frac{\pi(q-1)}{q} \tag{8}$$

Since $q \geq 2$, then argument of the first sine is restricted to the interval $(0, \pi)$. For the second sine consider this inequality:

$$-\frac{\pi}{q} < \theta < \frac{\pi}{q} \tag{9}$$

Multiplying it by -1 and adding the result to (8), we arrive at

$$0 < \frac{\pi l}{q} - \theta < \pi \tag{10}$$

which concludes the proof of the lemma. ∎

Lemma 2 If a weight vector $\mathbf{W}$ does not correctly classify $\mathbf{X}_j$, i.e. $\varepsilon_j \neq 0$, then

$$\mathrm{Re}(\overline{\varepsilon}_j \mathbf{W}^T \mathbf{X}_j) \leq 0 \tag{11}$$

Proof: Let $\mathbf{W}^T \mathbf{X}_j = a(\cos(\phi) + i\ \sin(\phi))$, with $a > 0$. Using (7) we can write

$$\mathrm{Re}(\overline{\varepsilon_j} \mathbf{W}^T \mathbf{X}_j) = 2a \sin(\frac{\pi l}{q}) \sin(\frac{\pi l}{q} - \phi) \tag{12}$$

As in the proof of the previous lemma, the first sine is always positive. To complete the proof of the lemma, we show that the second sine is non-positive. Taking ϕ to be in the interval $(0, 2\pi)$, we have the following inequality:

$$\frac{2\pi l}{q} - \frac{\pi}{q} \leq \phi \leq \frac{2\pi l}{q} + \frac{\pi}{q} \tag{13}$$

Multiplying this inequality by -1, we have

$$-\frac{2\pi l}{q} - \frac{\pi}{q} \leq -\phi \leq -\frac{2\pi l}{q} + \frac{\pi}{q} \tag{14}$$

Now we add $\frac{\pi l}{q}$ to all sides:

$$\frac{\pi l}{q} - \frac{2\pi l}{q} - \frac{\pi}{q} \leq \frac{\pi l}{q} - \phi \leq \frac{\pi l}{q} - \frac{2\pi l}{q} + \frac{\pi}{q} \tag{15}$$

Recalling that $l = 1, 2, \ldots, q-1$ and taking the extremes, we conclude that

$$-\pi \leq \frac{\pi l}{q} - \phi \leq 0 \tag{16}$$

This implies that $\sin(\frac{\pi l}{q} - \phi) \leq 0$, which concludes the proof of the lemma. ∎

Proof of Theorem 1: Without loss of generality we consider only the steps at which a weight change takes place ($\varepsilon \neq 0$) and $\alpha = 1$. Thus the weight vector at step k can be written as:

$$\mathbf{W}(k+1) = \mathbf{W}(1) + \varepsilon_1 \overline{\mathbf{X}}(1) + \varepsilon_2 \overline{\mathbf{X}}(2) + \ldots + \varepsilon_k \overline{\mathbf{X}}(k) \tag{17}$$

Now take the complex conjugate of both sides of the equation:

$$\overline{\mathbf{W}}(k+1) = \overline{\mathbf{W}}(1) + \overline{\varepsilon}_1 \mathbf{X}(1) + \overline{\varepsilon} \mathbf{X}(2) + \ldots + \overline{\varepsilon}_k \mathbf{X}(k) \tag{18}$$

Multiplying the equation by $\mathbf{W}^{*T}$ and taking the real parts of both sides we get:

$$\begin{aligned} \mathrm{Re}(\mathbf{W}^{*T} \overline{\mathbf{W}}(k+1)) \quad &= \quad \mathrm{Re}(\mathbf{W}^{*T} \overline{\mathbf{W}}(1)) + \mathrm{Re}(\overline{\varepsilon}_1 \mathbf{W}^{*T} \mathbf{X}(1)) + \\ & \quad\quad \mathrm{Re}(\overline{\varepsilon}_2 \mathbf{W}^{*T} \mathbf{X}(2)) + \ldots + \mathrm{Re}(\overline{\varepsilon}_k \mathbf{W}^{*T} \mathbf{X}(k)) \end{aligned} \tag{19}$$

Since $\mathbf{W}^*$ correctly classifies all input patterns, then from Lemma 1 we know that $\mathrm{Re}(\overline{\varepsilon}_j \mathbf{W}^{*T} \mathbf{X}(j)) > 0$, $j = 1, 2, \ldots, k$. Define a as follows:

$$a = \min_j(\mathrm{Re}(\overline{\varepsilon}_j \mathbf{W}^{*T} \mathbf{X}(j))) > 0 \tag{20}$$

The index j ranges over all possible ε_j and $\mathbf{X}(j)$, both of which are finite. From (19) we can write

$$\mathrm{Re}(\mathbf{W}^{*T} \overline{\mathbf{W}}(k+1)) \geq \mathrm{Re}(\mathbf{W}^{*T} \overline{\mathbf{W}}(1)) + ka \tag{21}$$

For all $z \in \mathbb{C}$ $|z| \geq \mathrm{Re}(z)$, and thus

$$\left| \mathbf{W}^{*T} \overline{\mathbf{W}}(k+1) \right| \geq \mathrm{Re}(\mathbf{W}^{*T} \overline{\mathbf{W}}(1)) + ka \tag{22}$$

The Cauchy-Schwartz inequality for two complex vectors $\mathbf{A}$ and $\mathbf{B}$ in a complex Hilbert space is $||\boldsymbol{A}||^2||\boldsymbol{B}||^2 \geq \left|(\boldsymbol{A}^T\overline{\boldsymbol{B}})\right|^2$ [Ber61, p. 30], and thus the following statement is true:

$$||\mathbf{W}(k+1)||^2 = ||\overline{\mathbf{W}}(k+1)||^2 \geq \frac{\left|\mathbf{W}^{*T}\overline{\mathbf{W}}(k+1)\right|^2}{||\mathbf{W}^*||^2} \tag{23}$$

For sufficiently large k the right side of (22) is positive and thus we can combine it with the above inequality:

$$||\mathbf{W}(k+1)||^2 \geq \frac{(\mathrm{Re}(\mathbf{W}^{*T}\overline{\mathbf{W}}(1)) + ka)^2}{||\mathbf{W}^*||^2} \tag{24}$$

Now we will derive another inequality which combined with Inequality (24), will lead to the conclusion that k is finite. We take the square of the norm of both sides of the update rule (5):

$$\begin{aligned}
||\mathbf{W}(j+1)||^2 &= (\mathbf{W}(j) + \varepsilon_j\overline{\mathbf{X}}(j))^T(\overline{\mathbf{W}(j) + \varepsilon_j\overline{\mathbf{X}}(j)}) \\
&= (\mathbf{W}^T(j) + \varepsilon_j\overline{\mathbf{X}}^T(j))(\overline{\mathbf{W}}(j) + \overline{\varepsilon}_j\mathbf{X}(j)) \\
&= ||\mathbf{W}(j)||^2 + ||\overline{\varepsilon}_j\mathbf{X}(j))||^2 + \overline{\varepsilon}_j\mathbf{W}^T(j)\mathbf{X}(j) + \varepsilon_j\overline{\mathbf{X}}^T(j)\overline{\mathbf{W}}(j) \\
&= ||\mathbf{W}(j)||^2 + ||\overline{\varepsilon}_j\mathbf{X}(j))||^2 + 2\mathrm{Re}(\overline{\varepsilon}_j\mathbf{W}^T(j)\mathbf{X}(j))
\end{aligned} \tag{25}$$

or

$$||\mathbf{W}(j+1)||^2 - ||\mathbf{W}(j)||^2 = ||\overline{\varepsilon}_j\mathbf{X}(j))||^2 + 2\mathrm{Re}(\overline{\varepsilon}_j\mathbf{W}^T(j)\mathbf{X}(j)) \tag{26}$$

Using Lemma 2 and letting $Q = \max_j ||\overline{\varepsilon}_j\mathbf{X}(j))||^2$, we get

$$||\mathbf{W}(j+1)||^2 - ||\mathbf{W}(j)||^2 \leq Q \tag{27}$$

We add the above inequalities for $j = 1, 2, \ldots, k$:

$$||\mathbf{W}(k+1)||^2 \leq ||\mathbf{W}(1)||^2 + Qk \tag{28}$$

Combining Inequalities (24) and (28) we conclude that

$$1 \geq \frac{(\mathrm{Re}(\mathbf{W}^{*T}\overline{\mathbf{W}}(1)) + ka)^2}{||\mathbf{W}^*||^2(||\mathbf{W}(1)||^2 + Qk)} \tag{29}$$

As k grows large the right side of the inequality grows without bound, which is impossible, and thus we conclude that k must be finite. ∎

Corollary 1 If the weight vector and input vectors are real valued and $q = 2$, then Theorem 1 becomes Rosenblatt's Perceptron Theorem. [Ros62].

4 The Continuous-valued Perceptron

The weight and input vectors of the Continuous Perceptron are again complex and the output s_i is defined by the following activation function:

$$s_i = \frac{h_i}{|h_i|} \tag{30}$$

where h_i is the net input.

$$h_i = \mathbf{W}^T\mathbf{X}_i \tag{31}$$

Thus, the outputs take continuous values on the the unit circle in the complex plane.

The error ε_j for input vector $\mathbf{X}_j$ is defined as

$$\varepsilon_j = \begin{cases} 0 & \text{if } |d_j - s_j| < \epsilon \\ d_j - s_j & \text{otherwise} \end{cases} \tag{32}$$

where ϵ is a tolerance parameter.

Theorem 2 Theorem 1 is valid for the Continuous Perceptron.

Lemmas 1 and 2 can be proven for the Continuous Perceptron as well. The corresponding lemmas follow.

Lemma 3 If a weight vector $\mathbf{W}$ correctly classifies a vector $\mathbf{X}_j$ and $\varepsilon \neq 0$ then

$$\mathrm{Re}(\overline{\varepsilon}\mathbf{W}^T\mathbf{X}_j) > 0 \tag{33}$$

Proof: Suppose that $\mathbf{W}$ correctly classifies $\mathbf{X}$. Then $\mathbf{W}^T\mathbf{X}_j = 1$. Let $\varepsilon = 1 - e^{i\theta}$ for some θ in the open interval $(0, 2\pi)$.

$$\mathrm{Re}(\overline{\varepsilon}\mathbf{W}^T\mathbf{X}) = \mathrm{Re}(\varepsilon) = 1 - \cos\theta > 0 \tag{34}$$

which concludes the proof. ∎

Lemma 4 If a weight vector $\mathbf{W}$ does not correctly classify $\mathbf{X}_j$, i.e. $\varepsilon_j < 0$, then

$$\mathrm{Re}(\overline{\varepsilon}_j\mathbf{W}^T\mathbf{X}_j) \leq 0 \tag{35}$$

Proof: Let $\mathbf{W}^T\mathbf{X}_j = a(\cos(\phi) + i\ \sin(\phi))$, with $a > 0$ and ϕ in the interval $(0, 2\pi)$. Then the error $\varepsilon_j = 1 - (\cos(\phi) + i\ \sin(\phi))$.

$$\begin{aligned}
\mathrm{Re}(\overline{\varepsilon}\mathbf{W}^T\mathbf{X}) &= \mathrm{Re}(\varepsilon)a\cos(\phi) + \mathrm{Im}(\varepsilon)a\sin(\phi) \\
&= a((1 - \cos(\phi))\cos(\phi) + \sin(\phi)\sin(\phi)) \\
&= a(\cos(\phi) - (\cos^2(\phi) + \sin^2(\phi))) \\
&= a(\cos(\phi) - 1) < 0
\end{aligned} \tag{36}$$

∎

We also define a as

$$a = \inf(\mathrm{Re}(\overline{\varepsilon}_j\mathbf{W}^{*T}\mathbf{X}(j))) > 0, \tag{37}$$

Had the error for the Continuous Perceptron been defined without the tolerance parameter ϵ (equation 32), a would have been 0.

Proof of Theorem 2: The proof of this theorem is identical to the proof of the q-state Perceptron with three changes:

1. Lemma 1 is replaced by Lemma 3.
2. Lemma 2 is replaced by Lemma 4.
3. Equation (20) is replaced by equation (37). ∎

Thus, the Continuous Perceptron can learn all continuous-valued functions it can represent, within an accuracy ϵ, in a finite number of steps.

5 Discussion and Conclusion

The Multivalued Perceptron can learn in finite time all functions $\mathbb{C}^n \rightarrow Q$, where Q is a discrete set of order q, that it can represent. Its learning capability is not merely the result of having $2n+2$ adjustable parameters (real weight components – two per complex weight) as opposed to the $n + 1$ adjustable parameters of the binary (Rosenblatt's) Perceptron: Two binary Perceptrons have $2n+2$ adjustable parameters, but together can only learn to classify vectors belonging to at most four classes; the Multivalue Perceptron will learn to classify vectors belonging to an arbitrary number of classes (q), as long as it is possible to do so.

The output of the Continuous Perceptron can be anywhere on the unit circle. The unit circle can in turn be mapped, using a continuous bijective function, on a real interval, such as $[-1, 1)$. Thus, we can view it as learning continuous-valued functions from $\mathbb{C}^n$ to $[-1, 1)$, within a given accuracy, in finite time. Clearly, the Continuous Perceptron is quite powerful.

The learning rule (5) looks very much like the complex LMS algorithm [WMB75]. In [Shy90] there is a discussion about the similarities and dissimilarities of the (real) LMS algorithm with the (real) binary perceptron. The Continuous Perceptron can also be used in the complex backpropagation algorithm [Geo92a].

At least part of the power of the new perceptrons should be attributed to the complex domain. The complex domain is being used in many disciplines to model physical processes: oscillators (electrical or mechanical), electrical circuits, in optics, in quantum mechanics, etc.. Being capable of capturing more complexity, the complex domain may yield more realistic biological neuron models, e.g. modeling phase differences in the arrival of signals at the neuron. An interesting application is in [Kak92], where the two dimensional nature of complex numbers is exploited to endow neurons with both local and global information. The complex domain is also promising in giving ultra-fast optical implementations [Noe88a, Noe88b]. Therefore, the complex domain promises to be a fertile ground for neural networks research.

Note: Most of the work reported here was done in [Geo92b].

Acknowledgement: The author wishes to thank Professor Subhash C. Kak of Louisiana State University for interesting discussions and suggestions.

References

[Ber61] Sterling K. Berberian. *Introduction to Hilbert Space*. Oxford University Press, New York, 1961.

[Geo92a] George M. Georgiou. Activation functions for neural networks in the complex domain. In *First International Conference on Fuzzy Theory and Technology*, Durham, N.C., October 1992.

[Geo92b] George M. Georgiou. *Parallel Distributed Processing in the Complex Domain*. PhD thesis, Tulane University, April 1992.

[GK92] George M. Georgiou and Cris Koutsougeras. Complex domain backpropagation. *IEEE Transactions on Circuits and Systems. II: Analog and Digital Signal Processing*, 39(5):330–334, May 1992.

[Kak92] Subhash C. Kak. State generators and complex neural memories. *Pranama - J. Phys.*, 38(3):271–278, March 1992.

[Noe88a] André J. Noest. Associative memory in sparse phasor neural networks. *Europhysics Letters*, 6(6):469–474, 1988.

[Noe88b] André J. Noest. Discrete-state phasor neural nets. *Physical Review A*, 38(4):2196–2199, 1988.

[Ros62] F. Rosenblatt. *Principles of Neurodynamics*. Spartan, New York, 1962.

[Shy90] J. J. Shynk. Performance of a single-layer perceptron. *IEEE Trans. on Neural Networks*, 1(3), 1990.

[WMB75] Bernard Widrow, John McCool, and Michael Ball. The complex LMS algorithm. *Proc. IEEE*, 63(4):719–720, 1975.