# SPECTRUM ESTIMATION BY NEURAL NETWORKS AND THEIR USE FOR TARGET CLASSIFICATION BY RADAR

Gerardo J. Melendez
US Army Communications & Electronics Command
Night Vision & Electronics Sensors Directorate
ATTN: AMSEL-RD-NV-RC
Fort Monmouth, NJ 07703
USA

Stanislav B. Kesler
Drexel University
Electrical & Computer Engineering Department
32-nd & Chestnut Streets
Philadelphia, PA 19104
USA

## ABSTRACT

Radar targets with moving components (engines, propellers, etc.) are well characterized in the frequency domain so that the estimation of spectral parameters can be used in the process of extracting features for their classification. With non-imaging scanning radars the application of spectral parameter estimation for target classification is limited by a short time on target and other radar parameters. To alleviate these limitations, artificial neural networks have been used bringing with them the long training time issue. This paper suggests a neural network approach that capitalizes on the prescence of spectral components and reduces the training time. This approach divides the training of the multi layer perceptron (MLP) in two steps: pretraining and final training. Pretraining guides the MLP to recognize the Fourier basis set of signals. The final training for target classification is then facilitated.

## 1. MOTIVATION & PREVIOUS WORK

This paper was motivated by the need to reduce spectral estimation limitations in the implementation of aircraft recognition algorithms using non-imaging radar returned signals. It has been observed that target signatures (modulations imposed on the returned RF energy produced by the moving components of the target) are well characterized in the frequency domain [1-2]. As a result, attempts have been made at developing target recognition algorithms to extract features in the frequency domain [3-4]. Since search and surveillance radar usually operate with scanning antennas, a short time on target (TOT) is a serious limitation for target recognition algorithms because the algorithms are usually designed as add-on processors on existing radars.

In order to alleviate the TOT and other limitations, an artificial neural networks (ANN) approach has been proposed [5]. This approach divides the training of the multi layer perceptron (MLP) in 2 steps: pretraining and final training. To implement pretraining, a basis set of simple signals is defined and an ANN trained to recognize that set of signals. Final training then consists of presenting data from the target set in question during which the network's training is facilitated by determining the presence of each basis set element present in the target's signature. If the Fourier basis set is selected for pretraining (an obvious choice for the radar application as discussed above), then pretraining becomes the problem of spectral estimation by ANNs.

Previous work in spectral estimation by ANN trained the network using the back propagation algorithm (BPA) [6]. There, the training data consisted of temporal signals for input and their corresponding magnitude of the DFT as the desired output.

## 2. APPROACH

The approach presented here (which is covered more extensively in [5]) constructs a network with each output node assigned to a frequency. If the input contains at least one of those frequencies, then the magnitude at the corresponding output node is required to be higher than that of the nodes whose frequencies are not present in the input.

Consider the input-output relationship of a 1 sigmoidal hidden layer network with a linear output layer with zero bias:

$$O^{(2)}(k) = \sum_{j=1}^{J} w_{kj}\, f\left( \sum_{i=1}^{I} w_{ji} x(i) + b_j \right). \tag{1}$$

Let each output node be assigned to a specific frequency, let the frequencies be equally spaced and let only one frequency to be present in the input signal. Assume the data is real thus we need only frequencies in the range $0 \le \omega < \pi$. Consider the entire frequency range by indexing the frequencies and input respectively

$$\omega_m = \frac{\pi(m-1)}{K}, \quad m = 1,\ldots,K \tag{2}$$

$$x_m(i) = \cos\omega_m i, \; i = 1,\ldots,I, \; I \ge 2K. \tag{3}$$

and the output is required to be

$$O_m^{(2)}(k) = \begin{cases} a & \text{for } m = k \\ b & \text{otherwise} \end{cases} \quad \text{where } a = \max_{\forall k} O_m^{(2)}(k). \tag{4}$$

Here we define $I$, $J$ and $K$ as the number of input, hidden and output nodes respectively. Next, define other network vectors and matrices

$$\mathbf{o}_m^{(2)} = \begin{bmatrix} O_m^{(2)}(1) \\ \vdots \\ O_m^{(2)}(K) \end{bmatrix}, \quad \mathbf{x}_m = \begin{bmatrix} x_m(1) \\ \vdots \\ x_m(I) \end{bmatrix},$$

$$\mathbf{W}^{(2)} = \left[ w_{kj} \right]_{\substack{k=1\ldots K \\ j=1\ldots J}} \quad \mathbf{W}^{(1)} = \left[ w_{ji} \right]_{\substack{j=1\ldots J \\ i=1\ldots I}} \quad \mathbf{b}^{(1)} = \left[ b_j \right]_{j=1\ldots J}$$

where $\mathbf{o}_m^{(2)}$ is $K$ by 1, $\mathbf{x}_m$ is $I$ by 1, $\mathbf{W}^{(2)}$ is $K$ by $J$, $\mathbf{W}^{(1)}$ is $J$ by $I$ and $\mathbf{b}^{(1)}$ is $J$ by 1. Then Eq. 1 can be expressed as

$$\mathbf{o}_m^{(2)} = \mathbf{W}^{(2)} f\left( \mathbf{W}^{(1)} \mathbf{x}_m + \mathbf{b}^{(1)} \right) \tag{5}$$


0-7803-2431-5/95 $4.00 © 1995 IEEE

where $f(x)$ with a matrix for operand implies that the function is applied to all of the elements in the matrix.

Next, construct the matrices for $K$ equations 5:

$$\mathbf{O}^{(2)} = \left[\mathbf{o}_1^{(2)} \cdots \mathbf{o}_K^{(2)}\right], \quad \mathbf{X} = \left[\mathbf{x}_1 \cdots \mathbf{x}_K\right] \tag{6}$$

where $\mathbf{O}^{(2)}$ is $K$ by $K$ and $\mathbf{X}$ is $I$ by $K$. In order to detect a sinusoid we want $\mathbf{O}_k^{(2)}$ to be such that $O_k^{(2)}(k) > O_k^{(2)}(p) \quad \forall p \neq k$. That is, the diagonal element in each column is greater than any other element in that column. We then say that $\mathbf{O}^{(2)}$ is diagonal-dominant with respect to its columns. We denote such a matrix as $\mathbf{D}$. Then from Eq. 5 we have

$$\mathbf{D} = \mathbf{W}^{(2)} \mathrm{f}\left(\mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)}\right). \tag{7}$$

Next we solve Eq. 7 for the weights and biases given the input $\mathbf{X}$. To achieve this assume the following two conditions

$$K = J \text{ and } \mathbf{W}^{(2)} = \mathbf{I}. \tag{8}$$

Then

$$\mathbf{D} = \mathrm{f}\left(\mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)}\right) \Rightarrow \mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)} = \mathrm{f}^{-1}(\mathbf{D}). \tag{9}$$

Note that $\mathrm{f}^{-1}(x)$ is monotonic increasing since $\mathrm{f}(x)$ also is monotonic increasing. Then we have that

$$\mathrm{f}^{-1}(\mathbf{D}) \equiv \hat{\mathbf{D}} \tag{10}$$

also is diagonal-dominant with respect to its columns. Now we need to solve $\mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)} = \hat{\mathbf{D}}$ when $\mathbf{X}$ is known. We claim that

$$\mathbf{W}^{(1)} = \mathbf{X}^T \text{ and } \mathbf{b}^{(1)} = [\mathbf{0}] \tag{11}$$

is a solution since $\mathbf{X}^T\mathbf{X}$ is diagonal-dominant with respect to its columns as required when Eqs. 10 and 11 are substituted in Eq. 9. To show that recall Eq. 6 and note that $\mathbf{X}^T\mathbf{X}$ is symmetric. Symmetry and the condition

$$\mathbf{x}_k \cdot \mathbf{x}_k > \mathbf{x}_l \cdot \mathbf{x}_k \quad \forall l \neq k. \tag{12}$$

imply that $\mathbf{X}^T\mathbf{X}$ is diagonal-dominant with respect to its columns and rows. Therefore, it suffices to show that Eq. 12 is true. Substitute Eq. 3 in Eq. 12 to obtain $\sum_{i=1}^{I} \cos^2 \omega_k i > \sum_{i=1}^{I} \cos \omega_k i \cos \omega_l i$ which can be shown to be true using trigonometric identities and the fact that $\sum_{i=1}^{I} \cos \omega_k i = 0$ if and only if the summation interval covers a complete number of cycles.

Two key issues that need to be addressed when implementing this approach are the facts that the development described above considered only zero phase inputs and that only one sinusoid per input sequence was allowed. Consider the input's phase first.

To develop an implementation that deals with non zero input phases we used 3 dimensional plots of input phase $(\phi)$, input frequency $(\omega)$ and hidden node activity $(a)$ or output $(o)$ for the ranges $-\pi \leq \phi \leq \pi$ and $0 \leq \omega \leq \pi$ and $\mathrm{f}(a) = o$. The phase and frequency for which the node was matched to are denoted by $\phi_m$ and $\omega_m$, respectively and are annotated in the figure's caption. The number of input nodes used in the simulation was $I$=225.

Figure 1 presents the activity of a hidden node with $\phi_m = 0$ and $\omega_m = \frac{1}{2}\pi$. As expected, the maximum output value occurs when the input is perfectly matched to the selected $\phi_m$ and $\omega_m$ as indicated by the hump in Figure 1a and by the white section in the center of Figure 1b. Irrespective of $\phi_m$, changes in $\omega$ have the desired effect of reducing the output of the node to nearly zero. Changes in $\phi$ when $\omega = \omega_m$ do not produce an output consistent with the desired results. Note that in Figure 1 the output produced by a sinusoid input with $\omega = \omega_m$ and $|\phi| \geq \frac{\pi}{2}$ is less than the output produced by a sinusoid at any $\omega \neq \omega_m$. This situation is resolved by constructing a network which adds the output from 3 nodes each matched to a different $\phi_m$ $\left(\text{i.e. } 0, \frac{2}{3}\pi, -\frac{2}{3}\pi\right)$ and with each node's input normalized and biased by $\nu = \frac{16}{I}$ and -4 respectively. Such a network is shown in Figure 2 with its output shown in Figure 3 for all $\phi$ and $\omega$.

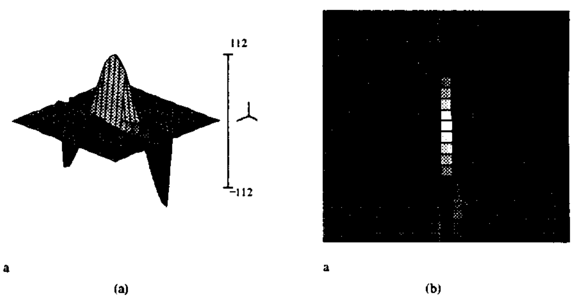


Figure 1. Activity of a node. Node has been matched to $\varphi_m = 0$ and $\omega_m = \frac{1}{2}\pi$.

The next issue to be addressed deals with the fact that more than one sinusoid could be present in the input sequence. The question to be answered is whether each individual output node will produce a higher value when its corresponding frequency is present regardless of whether some other frequencies are also present. This follows directly by letting the input sequence be $\mathbf{x} = \sum_{k=1}^{K} a_k \mathbf{x}_k$, where $a_k = 0$ or 1, substituting into the network's equation and noting that $\mathrm{f}(x)$ increases monotonically.

To apply this approach to the target classification problem, a network is constructed to estimate the anticipated required number of frequencies using the construction of Figure 2. Then a layer with $L$ nodes ($L$=number of classes) is appended with its weights, $\mathbf{W}^{(3)}$, randomly initialized. This produces a 2-hidden layer network with the 2-nd hidden layer consisting of linear nodes. This network is compressed into a 1-hidden layer network with a new weight matrix $\tilde{\mathbf{W}}^{(2)}$ and a new bias vector $\tilde{\mathbf{b}}^{(2)}$ given by

$$\tilde{\mathbf{W}}^{(2)} = \mathbf{W}^{(3)}\mathbf{W}^{(2)} \text{ and } \tilde{\mathbf{b}}^{(2)} = \mathbf{W}^{(3)}\mathbf{b}^{(2)} + \mathbf{b}^{(3)} \tag{13}$$

We call a network constructed in this fashion a pretrained network which can then be used to be trained (final training) against the $L$ different classes using the BPA.

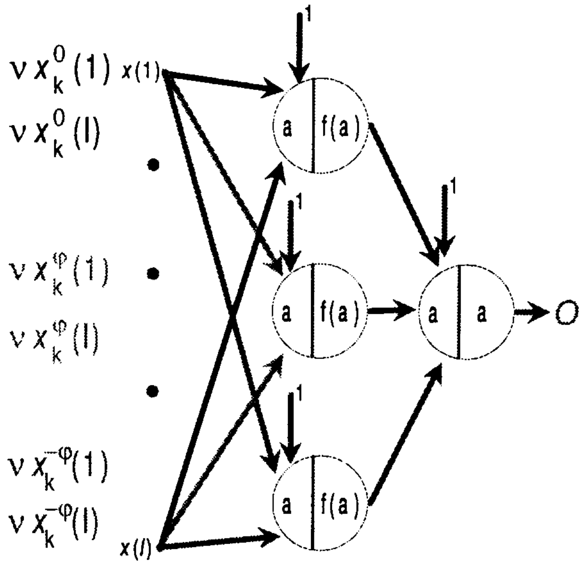


Figure 2. Network for sinusoid detection of a single frequency.

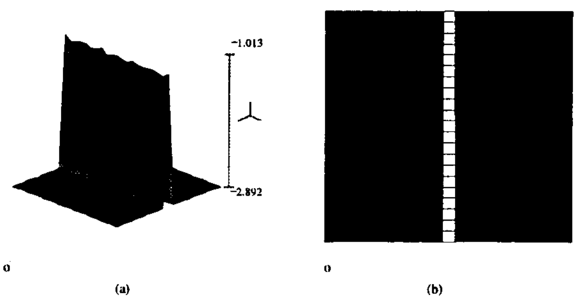


Figure 3. Addition of output from 3 nodes with normalized and biased activity. Each node has a different phase with values 0, $\frac{2}{3}\pi$ and $-\frac{2}{3}\pi$.

## 3. RESULTS

Two different sets of results are presented. First, the spectrum estimation capabilities of the constructed network are verified using synthetic data. Second, classification results for a 2-class classifier using actual radar data from two aircrafts are presented for a baseline approach and the pretrained approach.

### 3.1. Spectrum estimation verification

A network was constructed using the architecture of Fig. 2 with a (40, 60, 20) topology (i.e. 40 input nodes, 60 hidden nodes and 20 output nodes). The output produced by the signals:

$$x(i) = \cos(6\pi i/20) + 0.5\cos(11\pi i/20 + 0.5\pi) + 0.5\cos(14\pi i/20 - 0.5\pi) \quad (14)$$

$$x(i) = \cos(6.25\pi i/20) + \cos(11.5\pi i/20) + \cos(16.75\pi i/20) \quad (15)$$

are shown in Figures 4-5 respectively.

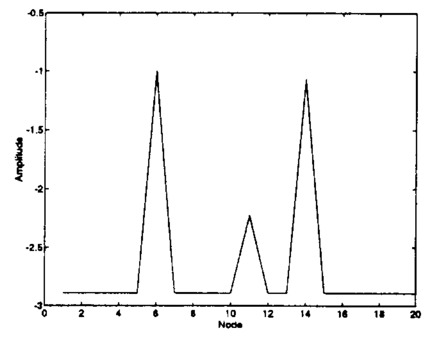


Figure 4. Output for multiple sinusoids with non zero phase.

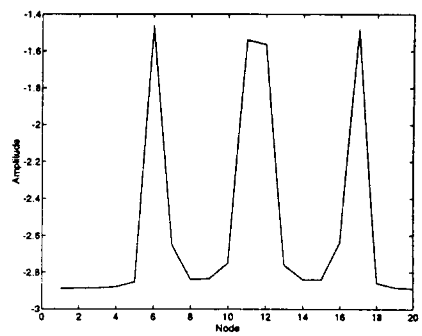


Figure 5. Output for multiple sinusoids with in-bin frequencies.

### 3.2. Target classification performance

In order to experimentally assess the effect of pretraining, two approaches were implemented: a baseline approach (BA) and a pretrained approach (PA).

The BA consisted of a 1 hidden layer network, with randomly initialized weights and trained using the BPA. Two experimental variables were considered: the learning rate $\eta$ and the number of hidden nodes $J$. Three $\eta$ levels were used: low ($\eta = 0.0001$), medium ($\eta = 0.001$) and high ($\eta = 0.01$ and $\eta = 0.02$). These values were experimentally obtained to establish a range in $\eta$ where the network went from incomplete training (low $\eta$) to oscillatory training (high $\eta$). In this way, the comparisons between the two approaches could be made without any bias on the selection of $\eta$. The selection of $J$ was initiated with the smallest nontrivial value ($J$=2) and allowed to increase until memorization was evident in the performance.

Since the only difference between the PA and BA is the method used to initialize the weights, the same experimental variables described above for the BA apply to the PA.

Two sources of signals were used. The training signals were obtained from a target signature generator which provided actual radar signals while allowing engine speed and signal to noise ratio to be varied. The test signals were obtained from an actual X-band, CW, FM tracking radar. All of the data was preprocessed to remove the skinline and normalize it to the target's radial velocity and range.

Tables 1 and 2 present the performance for the two approaches respectively using actual radar signals from two different aircrafts. The first column identifies each run by a number. The next two columns list the values for the experimental variables. The fourth column lists the best classification rate for the 10 attempts. The average column list the average classification rate over the 10 attempts. The last column indicates how many attempts in the run were successful. Success was defined as attaining an overall classification rate bigger than or equal to 80% while maintaining the lowest classification rate (i.e. for either class 1 or 2) at 75% or greater. For every combination of variables (each row), 10 different attempts were performed with each attempt consisting of a different random weight initialization (all weights for BA and the appended weights for PA). All attempts were limited to 200 epochs.

Table 1. Baseline Approach Performance

| Run | $J$ | $\eta$ | Best Rate | Average | Success |
|---|---|---|---|---|---|
| 1 | 2 | 0.0001 | 0.53 | 0.51 | 0(1) |
| 2 | 4 | 0.0001 | 0.53 | 0.51 | 0(1) |
| 3 | 6 | 0.0001 | 0.53 | 0.49 | 0(1) |
| 4 | 2 | 0.001 | 0.54 | 0.52 | 0 |
| 5 | 4 | 0.001 | 0.58 | 0.53 | 0 |
| 6 | 6 | 0.001 | 0.56 | 0.52 | 0 |
| 7 | 2 | 0.01 | 0.54 | 0.52 | 0 |
| 8 | 4 | 0.01 | 0.57 | 0.53 | 0 |
| 9 | 6 | 0.01 | 0.56 | 0.52 | 0 |
| 10 | 2 | 0.02 | 0.54 | 0.52 | 0(2) |
| 11 | 4 | 0.02 | 0.58 | 0.53 | 0 |
| 12 | 6 | 0.02 | 0.52 | 0.52 | 0 |

Table 2. Pretrained Approach Performance

| Run | $J$ | $\eta$ | Best Rate | Average | Success |
|---|---|---|---|---|---|
| 1 | 6 | 0.0001 | 0.50 | 0.50 | 0(1) |
| 2 | 9 | 0.0001 | 0.71 | 0.53 | 0(1) |
| 3 | 6 | 0.001 | 0.90 | 0.76 | 5 |
| 4 | 9 | 0.001 | 0.72 | 0.61 | 0 |
| 5 | 6 | 0.01 | 0.89 | 0.66 | 0(2) |
| 6 | 9 | 0.01 | 0.73 | 0.55 | 0(2) |

Notes: (1) Training incomplete
(2) Oscillatory behavior

## 4. CONCLUSIONS

Figures 4 and 5 indicate that the pretraining by construction of a sinusoid detector was successful at estimating the spectrum. Figure 4 indicates that the network was capable of handling non-sero phase inputs as required by the development. An interesting point is that the approach development completely ignored the input amplitude. Figure 4, however, indicates that some of that information was extracted. Figure 5 stressed the fact that the input signals do not always consist of sinusoids at the exact frequencies being estimated.

Table 1 indicates that the BA did not performed well the task of target classification. The final training performance results listed in Table 2 for the PA indicate that the PAgeneralization exceeded the performance of the BA. This is particularly true for run 3 which produced the only successful runs obtained for all the experiments. 5 of the 10 attempts in that run were successful.

The PA approach has a number of advantages over the BA. The PA training time was reduced in two ways. First it reduced the required number of epochs for the attempts to attain a prespecified generalization classification rate. It was observed, that 200 epochs sufficed to obtain successful attempts for PA but not for BA. Second, due to the quantization of the size of the hidden layer, the combinations required to search for the "best" set of experimental variables was reduced. We observe that the BA required 12 runs and the PA required 6 runs.

Another possible advantage which was not conclusively accounted for is that the PA appears to require less training data than the BA. For some problems (target classification included) that advantage is of great importance.

## 5. REFERENCES


1. JH Dunn, DD Howard, "Target noise," in Radar Handbook, M.I. Skolnik ed., McGraw-Hill, New York, NY, pp. 28.1-28.28, 1970.
2. GJ Melendez, A Guez, SB Kesler, "Target classification using radar data: a comparative study," Proceedings SPIE Intelligent Information Systems Conference, Orlando, FL, April 21-24, 1992 pp.789-799.
3. MR Bell, RA Grubs, "JEM modeling and measurement for radar target identification," IEEE Transactions on Aerospace and Electronic Systems, Vol. 29, No. 1, January 1993, pp.73-87.
4. GJ Melendez, W Hannibal, S Flanagan, SB Kesler, "New developments in the area of helicopter identification using radar data," Proceedings 2-nd Automatic Target Recognizer Systems and Technology Conference, Fort Belvoir, VA, March 17-18, 1992, pp. 101-118.
5. GJ Melendez, "Pretraining and shelving of neural networks by Fourier analysis and their use for target classification," Doctoral Dissertation, Drexel University, Philadelphia, PA, 1993.
6. C Wang, MA Wickert, C Wu, "Three layer neural networks for spectral estimation, " Proceedings ICASSP, Albuquerque, NM, April 3-6, 1990, pp.881-884.