

Direct Estimation of Noisy Sinusoids Using Abductive Networks

R. E. Abdel-Aal

Center for Applied Physical Sciences, Research Institute,
King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Abstract

Spectral estimation techniques have been used for many years. In many cases, their complexity warrants investigating machine-learning alternatives where intensive computations are required only during training, with actual estimation simplified and speeded up. This allows using simple portable apparatus for fast and automated estimation in real time. We propose using abductive network machine learning for estimating both the amplitude and frequency of a single sine wave in the presence of additive Gaussian noise. Models synthesized by training on 1000 representative simulated sinusoids were evaluated on 500 new cases. With no phase variations and a signal to noise ratio of 7 dB, average absolute percentage errors for the sinusoid amplitude and period are 8.4% and 3.6%, respectively. Effects of the range of frequency variations and the noise level on the complexity and accuracy of the models were investigated. Amplitude and period estimates show signs of biased at a signal to noise ratio of 3 dB. Error variances track the Cramer-Rao bounds at high noise levels, with no thresholding observed down to 0 dB. The method is compared with a neural network model and with conventional DFT (discrete Fourier transform) based techniques and a Prony's based approach. The new approach is particularly useful when only a small portion of the sinusoid cycle is measured.

Index Terms: Spectral analysis, Frequency estimation, Parameter estimation, Machine learning, Abductive networks, Cramer-Rao bounds, Sinusoid, Gaussian noise.

Dr. R. E. Abdel-Aal,
P. O. Box 1759,
KFUPM, Dhahran 31261
Saudi Arabia

Phone: +966 3 860 4320, Fax: +966 3 860 4281
e-mail: radwan@kfupm.edu.sa

1. Introduction

Estimation of unknown parameters of sinusoids in noise is important in many areas, including underwater acoustics, radar, sonar, radio direction finding, and communication systems. Example applications include: determining carrier frequencies and baud rates of communication systems following distortion and nonlinearities introduced through transmission, analysis of earth waves, processing of nuclear magnetic resonance (NMR) signals, speech and image processing, modal analysis, condition monitoring for engineering structures, and control and protection of electrical power systems. The more generalized problem of spectral estimation has received considerable attention in the literature, see Kay and Marple Jr. (1981) for an overview. Although the fast Fourier transform (FFT) approach is computationally efficient and produces acceptable results in many situations, its frequency resolution in Hz is limited to the reciprocal of the width of the measurement time window. The method also suffers from spectrum smearing due to the leakage associated with the implicit windowing with a boxcar function. With pulsed signals, e.g. in radar, the resulting resolution may not be adequate, and leakage has a detrimental effect on the detectability of sinusoidal components. Many alternative procedures have been proposed to alleviate the above limitations (Marple Jr., 1989). Real-time spectral estimation is important in several areas; e.g. speech processing (Moreno and Fonollosa, 1992), communications (Sills and Black, 1996), and online monitoring, control, and protection of power systems (Osowski, 1992). However, many of the alternative spectral estimation techniques require intensive matrix computations and/or iterative optimization which may not be practical for real time processing and may not even converge to a solution (Marple Jr., 1989). This has often led to adopting sub-optimal alternatives to reduce the computational load to a manageable level. Some techniques may require user intervention. For example, the most appropriate order for the auto regressive (AR) model is usually not known a priori, making it necessary to explore several possible orders based on the minimization of an error criterion. The model order may be progressively increased until the criterion for the error power reaches a

minimum. However, for least squares estimation procedures, prediction error decreases monotonically with increased model order, and some penalty function may be required to discourage the selection of too high an order (Marple Jr., 1989). All the computational complexities have to be repeated for each individual spectrum estimation, even though it may be quite similar to a previous one. These factors are not conducive to speeding up and automating spectrum analysis for real-time applications.

A recent trend in many areas of applied sciences has been to resort to a machine learning approach when a rigorous algorithmic solution becomes too complex. With this approach, a model is automatically developed through training on an adequate number of solved examples. Once the model is synthesized, it can be used to perform fast predictions of outputs corresponding to new cases. A major advantage of this approach in spectral analysis is that intensive computations are now required only once, i.e. during training for model synthesis, rather than being repeated for every sample analyzed during actual use. Using the model to process new data becomes a simple and speedy operation that can be implemented in real time using compact and portable apparatus. While conventional spectral estimation techniques are primarily concerned with frequency estimates only, with amplitudes and phase requiring extra computations, a machine learning solution can estimate all parameters of interest simultaneously.

Many techniques exist for the development of machine learning systems (Weiss and Kulikowski, 1991), including statistical pattern recognition methods (e.g. Bayesian classifiers and discriminant functions), artificial neural networks, and methods for the induction of decision trees. These techniques vary in their accuracy, complexity, computational requirements during training, and their ability to provide human-like explanations for their conclusions. Such variations have led to newer techniques that combine good features from various methods. An example of such 'hybrids' is the AIM (abductive induction mechanism) modeling tool

(AbTech Corporation, 1990; Montgomery and Drake, 1990) which draws on statistical and multiple regression analysis methods as well as neural networks, resulting in a faster and more automated approach to model synthesis. The development of this approach of machine learning through self-organization has followed the track of the group method of data handling (GMDH) algorithm (Farlow, 1984) and the closely related adaptive learning network (ALN) technique (Barron et al., 1984). A mathematical description of the GMDH foundations for AIM is given in (Abdel-Aal, 1998).

Neural network techniques have been proposed for signal identification and parameter estimation for multiple sinusoids in noise. Analog neural networks have been proposed to solve optimization differential equations using the gradient descent method in real time, mainly for waveform analysis in power systems, e.g. (Osowski, 1992). These are implementations of the Hopfield type feedback recurrent dynamic networks that do not require a separate training phase. An analog processing element is needed for each sample in the time record processed, and therefore implementation cost would be high for large waveform records. To avoid convergence to a local minimum of the network energy function, the network may have to be initialized several times with different initial values in search of the global minimum. In another application of more conventional neural networks, nonlinear constrained Hebbian learning algorithms were developed for frequency estimation of multiple sinusoids in impulsive and colored noise (Karhunen and Joutsensalo, 1992). A radial basis function (RBF) neural network has been described for estimating the input noise density function and its derivative, which are fed to a locally optimum detection procedure for determining weak sinusoidal signals in non Gaussian noise (Hummels et al., 1995). An analysis-by-synthesis system of neural networks was used to extract the sinusoidal parameters from the spectra of audio signals (Garcia, 2001).

This paper investigates the use of AIM machine learning for estimating the amplitude and frequency of a single sinusoid in Gaussian noise. Parameter estimation of single sinusoids has

been reported using various techniques (Hancke, 1990; Nishiyama, 1997; Kay, S., 1989; Nishiyama, K., 1999; and Rife and Boorstyn, 1974). Previous experience with AIM in modeling and forecasting the daily minimum temperature (Abdel-Aal and Elhadidy, 1994) has indicated improved prediction accuracy and faster training compared to neural network models. Model synthesis is also more automated with AIM, requiring little or no user intervention. Expressed as a network of polynomials, AIM models can provide better insight into the phenomenon being modeled. The technique has also been used for estimating the position, width, and height of single and double noisy Gaussian peaks in nuclear spectra (Abdel-Aal, 1998). Following a brief overview of AIM, models for estimating both noiseless and noisy sinusoids are synthesized and evaluated, and the results compared with the corresponding Cramer-Rao bounds and other conventional and non-conventional spectral estimation methods.

2. AIM Abductive Machine Learning

AIM is a supervised inductive machine-learning tool for automatically synthesizing abductive network models from a database of input and output values which represent a training set of example situations. Once synthesized by training on a training data set, the network can be queried with new input data to provide the corresponding predicted output. Abductive networks (Montgomery and Drake, 1990) combine the advantages of the neural network approach with those of advanced statistical methods. While the processing elements in neural networks are restricted by the neuron analogy, AIM builds networks of various types of more powerful numerical functional elements based on prediction performance. The network size, element types, connectivity, and coefficients for the optimum model are automatically determined using well-proven optimization criteria, thus reducing the need for user intervention. With neural networks, the user has to experiment with various architectures and there are no hard and fast design rules to determine optimum values for the number of hidden layers, number of neurons in each hidden layer, and various training parameters. Often a number of combinations need to be

tried in search of the best solution. With the commonly used standard back propagation algorithm, training times can be huge and there are many training parameters to adjust which may have a major effect on the results (Weiss and Kulikowski, 1991). The algorithm is not guaranteed to converge to a good solution, and because the method may be unstable and oscillate between solutions, it may not be clear when to stop (Weiss and Kulikowski, 1991). This makes AIM much easier to use and considerably reduces the learning/development time and effort.

AIM models take the form of layered feed-forward abductive networks of functional elements (nodes) (AbTech Corporation, 1990), see Fig.1. Elements in the first layer operate on various combinations of the independent input variables (x's) and the single element in the final layer produces the predicted output for the dependent variable y. In addition to the functional elements in the main layers of the network, an input layer of normalizers converts the input variables into an internal representation as Z scores with zero mean and unity variance, and an output layer of unitizers restores the results to the original problem space. Both the element type and the combination of inputs to it from all the previous layers are selected automatically for best prediction performance according to the predicted squared error (PSE) criterion (Barron, 1984). The used version of AIM supports the following main elements:

(i) A white element which consists of a constant plus the linear weighted sum of all outputs of the previous layer, i.e.:

$$\text{"White" Output} = W_0 + W_1x_1 + W_2x_2 + W_3x_3 + \dots + W_nx_n \quad (1)$$

where x_1, x_2, \dots, x_n are the inputs to the element and W_0, W_1, \dots, W_n are the element weights.

(ii) Single, double, and triple elements which implement a third-degree polynomial expression with all possible cross-terms for one, two, and three inputs respectively; for example,

$$\text{"Double" Output} = W_0 + W_1x_1 + W_2x_2 + W_3x_1^2 + W_4x_2^2 + W_5x_1x_2 + W_6x_1^3 + W_7x_2^3 \quad (2)$$

The first step in solving a problem is preparing a database of input-output solved training examples, which AIM uses to synthesize the model network layer by layer until no further improvement in performance is possible or a preset limit on the number of layers is reached. Within each layer, every element is computed and its performance scored for all combinations of allowed inputs. The best network structure, element types and coefficients, and connectivity are all determined automatically by minimizing the PSE criterion. This selects the most accurate model that does not overfit the training data, and therefore strikes a balance between the accuracy of the model in representing the training data and its generality which allows it to fit yet unseen future data. In this way the model is optimized for the actual use for which it is developed, rather than only at the training phase. The user may optionally control this trade-off between accuracy and generality using the complexity penalty multiplier (CPM) parameter (AbTech Corporation, 1990). Larger values than the default value of 1 lead to simpler models that are less accurate but are more likely to generalize well with unseen data, while lower values produce more complex networks that may overfit the training data and degrade prediction performance with noise. To obtain good AIM models, the training set should be a good representation of the problem space. AIM's learning task is also simplified by breaking the problem into smaller and more manageable assignments, and by utilizing knowledge on relevant parameters in the choice of input variables for the training database.

3. Estimating Noiseless Sinusoids

As a first exercise, AIM was trained on noiseless sinusoids of various amplitudes and frequencies, with the objective of developing a model that determines the amplitude and period of the sine wave automatically. The model was synthesized by training on 1000 simulated sinusoids having the same phase angle of zero but randomly generated values for both the amplitude and the period. The data record for a noiseless sinusoid having amplitude A and period T was represented as:

$$y(i) = A \sin \frac{2\pi i}{T}; i = 1, 2, \dots, 50 \quad (3)$$

Samples are uniformly spaced in time and the sample spacing is taken as 1 second for simplicity. Initially, both amplitude and period variations ranged over a decade, i.e. a ratio of ten between the maximum and minimum values. The amplitude A assumed real values which are uniformly distributed between 1.0 and 10.0, inclusive ($1.0 \leq A \leq 10.0$), while the period T assumed integer values which are uniformly distributed between 5 and 50, inclusive ($5 \leq T \leq 50$). Uniform distribution for A and T values allowed equal representation of different sinusoids during training for model synthesis. A sinusoid consisted of 50 samples that constituted the input variables to AIM. Restriction on the number of samples in a record was imposed by the maximum limit of 50 input variables for the AIM version used. With $5 \leq T \leq 50$, the largest period (lowest frequency) represents one full cycle of the sinusoid within the time data record input to AIM, while the shortest period (highest frequency) contains five samples within each sinusoid cycle, which adequately satisfies the sampling theorem.

Records for the AIM training/evaluation database were derived by appending corresponding known values for the amplitude and frequency to each sinusoidal data records computed in (3).

A typical complete AIM record is represented below:

Inputs :	Outputs :
Values of the Sinusoid Samples	Corresponding Sinusoid Parameters
$y(1) \ y(2) \ y(3) \ \dots \ y(50)$	Amplitude, A Period, T

AIM generates a model for each variable declared as output in the training database. All model synthesis was performed using the default value for the complexity penalty multiplier (CPM=1). The top row in Fig. 2 shows the abductive network models synthesized for the amplitude (left) and the period (right). Here y_i indicates $y(i)$, the i th time sample of the sinusoid. Multiple input samples may be indicated on the same model input line to reduce figure complexity. The

amplitude model is a relatively simple 1-layer, 3-input network while the period model is a more complex 4-layer, 9-input model. Determining the amplitude of a noiseless sinusoid is simpler than determining its frequency.

To investigate how model complexity is affected by the range of frequency variations in the training data, training was repeated with variations in the waveform period limited to only one octave, i.e. a ratio of two between the maximum and minimum periods. Three ranges were used for the frequency octave variations: a high frequency range ($5 \leq T \leq 10$), a medium frequency range ($25 \leq T \leq 50$), and a low frequency range ($100 \leq T \leq 200$). Resulting models for these frequency octave ranges are shown in the remaining three rows of Fig. 2. All models for the octave frequency variations are clearly simpler than their decade counterpart, for both the amplitude and the period. It is noted that the 'white' functional element is only a linear weighted sum of its inputs, and therefore is simpler than the 'triple' element that can contain terms of the third degree of the inputs. The three ranges for octave frequency variations have nearly the same complexity for the period models. However, the amplitude model is simplest for the medium frequency range, where the model is just a linear weighted sum of only two samples. Accurate amplitude estimation becomes relatively more difficult (requiring a more complex model) at both lower frequencies (shallower waveforms) and higher frequencies (fewer samples within the waveform period). This suggests that the medium frequency models should give the best overall estimation performance.

Fig. 3 shows a detailed description of the simple amplitude model obtained for medium frequency octave variations ($1.0 \leq A \leq 10.0$ and $25 \leq T \leq 50$). Equations shown are the outcome of model synthesis. Symbolically substituting for the equations and simplifying, we obtain the following relationship for the estimate of A in terms of two sinusoid samples:

$$\hat{A} = 0.027 + 1.147 y(8) + 0.155 y(24) \quad (4)$$

For example, the sinusoid $y(i) = 7.5 \sin(2\pi i / 40)$ has $A = 7.5$, $T = 40$, $y(8) = 7.133$, and $y(24) = -4.408$. Substituting for $y(8)$ and $y(24)$, (4) gives the estimate \hat{A} as 7.525, which is the correct sinusoid amplitude with an error of 0.33%.

To assess the performance of the models over a wide range of test cases, the AIM 'Evaluate' utility was used to run the resulting models through new evaluation sets (previously unseen during model synthesis) of 500 sinusoid records each. AIM predictions for the amplitude and period of each sinusoid were compared with the corresponding known values. Table 1 summarizes data on the percentage of the evaluation population having percentage errors within $\pm 2\%$, $\pm 5\%$, and $\pm 10\%$ of the true values of the sinusoid amplitude and period. Included also are the values of the overall mean and standard deviation (SD) of the absolute percentage error over the total evaluation population. In general, errors are markedly larger for the decade variations as compared to the octave variations. With the wider range of variations in the decade case, there is a fewer number of training examples in each narrow bin containing a set of close frequencies and amplitudes. With these bins having the same width in both cases, their number would be considerably larger in the decade case. Using the same total number of 1000 training examples in both cases, fewer number of the uniformly distributed, randomly generated training examples would fall within each bin, which leads to poorer learning. This problem can be solved by proportionally increasing the number of training examples for the case of decade variations.

For all octave ranges, amplitude and period estimates have an average absolute percentage error below 1.6% and the maximum error magnitude hardly exceeds 10%. As expected, models for the medium frequency octave give the best overall performance. It is noted that for the low frequency octave, the period is estimated more accurately than the amplitude. With the 50-sample time records used, conventional FFT based frequency estimation has a resolution of 0.02 Hz. For the low frequency octave ($100 \leq T \leq 200$), the measured record of 50 samples contains one half to one quarter of the sinusoid cycle, respectively, and frequencies fall within the range

0.005 Hz to 0.01 Hz. These two frequency limits are below the FFT resolution, and therefore conventional methods would most likely estimate all frequencies of this octave as the DC component, which amounts to a frequency estimation error of 100%. This shows that the proposed technique is particularly useful for frequency estimation from short time records that represent only a small portion of one cycle of the measured sinusoid. Further investigations reported here will consider only the case of the medium range frequency octave ($1.0 \leq A \leq 10.0$ and $25 \leq T \leq 50$).

4. Estimating Noisy Sinusoids

Realistic observation data often have the true sinusoid corrupted by broadband background and sensor noise. A noisy sinusoid may be represented as:

$$y(i) = A \sin \frac{2\pi i}{T} + e(i); i = 1, 2, \dots, 50 \quad (5)$$

where $e(i)$ are samples of a white Gaussian noise process with zero mean and standard deviation σ . The value of σ is chosen to give the required signal to noise ratio (SNR) given in decibels by:

$$SNR = 10 \log \frac{A^2}{2\sigma^2} \quad (6)$$

Model training and evaluation was repeated using signals for the medium range frequency octave at various noise levels. Fig. 4 shows the models synthesized with $SNR = 17$ and 3 dB, in comparison with the models derived in Section 3 above for noiseless data. The figure shows a general trend of increasing model complexity with increased noise level. This is expected as the model now tries to reconcile greater variations in the input data to the same true value for the output parameter. The most complex network is that for estimating the period at the largest noise level ($SNR = 3$ dB). Amplitude models remained simpler than period models for the noise levels considered. Each model was evaluated using 500 sinusoid records at the same noise level used for training, and statistics on the error are given in Table 2. As expected, the trend of increasing error with increased noise is obvious. It is clear that for noisy signals, frequency

estimates are more accurate than amplitude estimates. At $SNR = 3$ dB, 91.6% of the evaluation population has a period error within $\pm 10\%$, and the overall mean absolute period error is about 4.5%.

To highlight the increase in model complexity attributed to noise, Fig. 5 shows the detailed structure and equations for the amplitude model for $SNR = 17$ dB (left hand side of the middle row in Fig. 4). Fig. 5 should be compared with Fig. 3 which shows the 2-input model obtained for the noiseless case. Both models are 1-element linear models, but accounting for the noise has increased the number of inputs used to 23. Symbolic substitution for the equations in Fig. 5 and simplifying gives the following relationship for the amplitude estimate in terms of 23 samples of the noisy sine wave:

$$\begin{aligned} \hat{A} = & 0.035 + 0.075 y(2) + 0.059 y(3) + 0.087 y(4) + 0.093 y(5) + 0.121 y(6) + 0.141 y(7) \\ & + 0.124 y(8) + 0.127 y(9) + 0.078 y(10) + 0.102 y(11) + 0.086 y(12) + 0.044 y(13) \\ & + 0.053 y(14) + 0.026 y(15) - 0.040 y(18) - 0.060 y(22) - 0.026 y(24) - 0.033 y(25) \\ & + 0.020 y(33) - 0.042 y(37) - 0.031 y(39) - 0.036 y(43) - 0.029 y(46) \end{aligned} \quad (7)$$

Substituting the relevant waveform samples for the noiseless sinusoid having $A = 7.5$ and $T = 40$ into this equation gives $\hat{A} = 7.579$, which is accurate to 1.05%. Equation (7) should be compared with the simple 3-term expression in equation (4) for the noiseless case. Higher levels of noise would introduce even greater model complexity and nonlinearity. For example, at $SNR = 3$ dB, the amplitude model (left hand side of the bottom row in Fig. 4) becomes a 4-layer model with three nonlinear ‘Triple’ elements.

We have so far considered the absolute value of the percentage error as a measure of the estimation accuracy. Many frequency estimators suffer from being biased at low signal to noise ratios when the mean estimation error (not its absolute value) over the entire population becomes significantly different from zero. A positive mean value shows that the method tends to

overestimate, while a negative value is a sign of underestimation. In practice, we can only infer about performance of the estimator on the entire error population from measured performance on evaluation samples of finite size. Values of the mean error with such samples are assumed to be normally distributed about the true population mean for the error, with a standard deviation equal to the population standard deviation/ \sqrt{n} , where n is the sample size (Mendenhall and Beaver, 1994). For large values of n , (e.g. $n = 500$), we can also assume that the unknown standard deviation for the entire error population is equal to that of the finite evaluation sample. Judging whether the estimator is actually biased or not should take into account the statistical fluctuations in estimates of the mean error for such evaluation samples about the true mean error. The z statistic can be used to test the hypothesis that the true mean error is zero (i.e. the estimator is unbiased) at a specified confidence level, given the measured mean error for a finite evaluation sample. Table 3 shows data on the mean and standard deviation of the percentage error for the sinusoid amplitude and period as calculated from 500-case evaluation samples at two noise levels, $SNR = 7$ dB and 3 dB. At 95% confidence level ($\alpha = 0.05$), the hypothesis is accepted for $-1.96 < z < 1.96$. Data for both the amplitude and the period show that the hypothesis of unbiased estimation is rejected only for $SNR = 3$ dB. Therefore, the AIM estimators described are unbiased down to an SNR value of 7 dB, showing signs of bias at $SNR = 3$ dB.

So far, models were evaluated using evaluation data having the same noise level as that of the training data used to synthesize the model. Performance of models with evaluation data having different noise levels from that used to synthesis the model has also been investigated, and the results are shown in Table 4. Models developed with noiseless, $SNR = 17$ dB, and $SNR = 7$ dB training data were each evaluated with noiseless, $SNR = 17$ dB, and $SNR = 7$ dB evaluation data. The table shows that a model performs well at noise levels that are equal to or less than the noise level of the training data used to synthesize it. For example, models developed using training

data having $SNR = 7$ dB perform adequately with sinusoids having signal to noise ratios down to 7 dB. This ensures that the training data set gives an adequate representation of new data that may be encountered during model evaluation.

5. Comparison with Neural Networks

Performance of the abductive network model shown in the middle row of Fig. 4 for noisy sinusoids ($SNR = 17$ dB) was compared with a back propagation neural network model trained and evaluated on the same data used by the abductive model. The neural model was the default function approximation model synthesized by the NeuroExpert module of the NeuroSolutions 4 software for Windows. 20% of the training data were used for cross validation. The 3-layer network uses 50 neurons for the input layer, three neurons with a hyperbolic tangent transfer function for the hidden layer, and two output neurons representing the amplitude and the period of the sine wave. Combined together, the two AIM models for the amplitude and period use only 28 different samples of the sine wave, while the neural model requires all 50 inputs. Excluding irrelevant inputs helps simplify the resulting model and speed up its execution. It also reduces the effects of noise or measurement errors associated with such inputs. In contrast with the opaque nature of the neural models, the abductive models are quite transparent as evidenced by Fig. 5 and Equation (7) that describe the amplitude model. Expressing the model in an analytical form gives better insight into the modeled relationship, allows comparison with previously obtained forms for the relationship, and simplifies implementing the model and transporting it to other platforms. Table 5 compares the estimation performance of the abductive and neural models when evaluated on 500 sinusoid records at the same noise level used for training. Similar to the abductive model, the neural model estimates the period of the sine wave more accurately than the amplitude. The abductive model significantly outperforms the neural model, with the mean absolute error being 1.70% and 2.53%, respectively for the period, and 3.04% and 4.40%, respectively for the amplitude.

6. Comparison With the Cramer-Rao Bounds and Other Estimators

Performance of estimators is often compared against the Cramer-Rao (CR) bound (Norton, 1986) which sets a theoretical lower limit (a minimum variance unbiased bound) on the uncertainty that can be achieved in the estimation of a parameter from noisy measurements. A single complex sinusoid of N measured samples takes the form:

$$Y(i) = Ae^{j(2\pi Fi + \phi)} + v(i); i = 1, 2, \dots, N \quad (8)$$

where A is the amplitude, F the frequency, ϕ the phase, and $v(i)$ is a complex white Gaussian noise process with zero mean and variance $\sigma_v^2 = \sigma_{vr}^2 + \sigma_{vi}^2$. We assume that noise variances for the real and imaginary parts are equal, i.e. $\sigma_{vr}^2 = \sigma_{vi}^2$. For this simple case the Fisher's information J matrix (Norton, 1986) is a 3 x 3 matrix that can be easily inverted analytically to produce simple expressions for the bounds. For multiple sinusoids, the resulting larger J matrix is usually inverted by the computer and no attempt is made to derive analytical expressions for the bounds. If \hat{F} and \hat{A} are the estimates for F and A , then the Cramer-Rao bounds give the minimum variances on these estimates as (Nishiyama, 1997):

$$\min_var(\hat{F}) = \frac{\sigma_{vr}^2 p_0}{4\pi^2 A^2 (p_2 p_0 - p_1^2)} \quad (9), \text{ and}$$

$$\min_var(\hat{A}) = \frac{\sigma_{vr}^2}{p_0} \quad (10)$$

where p_j ; $j = 0, 1, 2$ are given by:

$$p_j = \sum_{i=1}^N i^j \quad (11)$$

$$\text{i.e. } p_0 = N, \quad p_1 = \frac{N(N+1)}{2}, \quad \text{and } p_2 = \frac{N(N+1)(2N+1)}{6} \quad (12)$$

The Cramer-Rao lower bounds (CRLB) on the frequency estimate is usually expressed in dB as:

$$\begin{aligned}
CRLB_F &= 10 \log \frac{1}{\min_var(\hat{F})} \\
&= 10 \log \frac{A^2}{2\sigma_{vr}^2} + 10 \log \left[8\pi^2 \left(\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4} \right) \right] \\
&= SNR + 10 \log \left[\frac{2\pi^2 N(N^2-1)}{3} \right]
\end{aligned} \tag{13}$$

where SNR is the signal to noise ratio for the sine/cosine part of the complex sinusoid as defined in (6). It is noted that the bound is independent of the frequency of the complex sinusoid, and is solely determined by N , the number of samples in the data record measured, and the signal-to-noise ratio. The bound is larger (error variance smaller) for larger SNR and N . For $N = 50$ we have:

$$CRLB_F = SNR + 59.149 \tag{14}$$

Substituting from (12) into (10), the CRLB bound on the amplitude estimate is:

$$\begin{aligned}
CRLB_A &= 10 \log \frac{1}{\min_var(\hat{A})} = 10 \log \frac{N}{\sigma_{vr}^2} \\
&= 10 \log \frac{A^2}{2\sigma_{vr}^2} + 10 \log \frac{2N}{A^2} \\
&= SNR + 10 \log \frac{2N}{A^2}
\end{aligned} \tag{15}$$

In addition to SNR and N , this CRLB depends on the amplitude of the sinusoid. For $N = 50$ and $A = 5$ we get:

$$CRLB_A = SNR + 6.021 \tag{16}$$

Fig. 6(a) and 6(b) plot $CRLB_F$ and $CRLB_A$, respectively, versus SNR for $N = 50$ and $A = 5$.

Although CRLB bounds for a single complex sinusoid have simple analytical expressions due to the simplicity of the J matrix, they do not generally apply to the corresponding real (cosine) or imaginary (sine) sinusoidal signal assumed here. A single real or imaginary sinusoidal signal at frequency F is composed of two complex sinusoids at frequencies $-F$ and $+F$. For these two

complex sinusoids, the CRLB bounds approach those of a single complex sinusoid when the frequency separation between the two sinusoids exceeds a critical frequency of $2/N$, i.e. $F > 1/N$ (Rife and Boorstyn, 1976). With $N = 50$ used here, this condition is satisfied for $F > 0.02$ Hz. With $N = 24$ used in (Tufts and Fiore, 1996), bounds for the single complex sinusoid were taken to represent those for a single real (cosine) wave having $F = 0.125$, i.e. $F = 3/N$.

Fig. 6 shows also plots of $10\log(1/MSE)$, where MSE is the mean square error in estimating the frequency F ($=1/T$) and amplitude A of the sine wave, versus the signal to noise ratio, SNR. Results were obtained when medium frequency AIM models ($1.0 \leq A \leq 10.0$ and $25 \leq T \leq 50$) synthesized at a given SNR were evaluated using 500 cases, all having the same nominal values of $A = 5$ and $T = 26$ ($F \approx 0.04 \approx 2/N$) at the same SNR value. Fig. 6(a) shows that the loss in frequency estimation accuracy by AIM from the CRLB improves (decreases) with the decrease in the signal-to-noise ratio and there are no signs of thresholding (sudden drop in the estimation accuracy) for SNRs as low as 0 dB. At $SNR = 0$, AIM's loss compared to $CRLB_F$ is less than 5 dB. At low signal to noise ratios, many frequency estimators exhibit a thresholding effect. For example, estimators employing iterative minimization, e.g. using the Gauss-Newton algorithm, suffer from the presence of local minima in the loss function to be minimized. At high noise levels, initial estimates will not be sufficiently accurate, and the algorithm may converge to a false local minimum, thus leading to a threshold SNR which must be exceeded for acceptable performance with noise (Stoica et al., 1989). Typical threshold SNR values where estimation accuracy for the frequencies of complex sinusoids departs sharply from tracking the CRLB are 7 dB for the principal component auto regressive method (PC-AR) (Rao and Raghavan, 1989) and -1 dB for the maximum likelihood (MLE) estimator (Kay, 1989).

Fig. 6(a) shows also the performance of another frequency estimator based on the Prony's method (Tufts and Fiore, 1996) for a single real sinusoid of the cosine form. Using a set of difference equations, the method reduces the problem of estimating the frequency of a cosine

wave to that of a linear least squares fit, giving the frequency estimate as (Tufts and Fiore, 1996):

$$\hat{F} = \frac{1}{2\pi} \cos^{-1} \left[\frac{\sum_{i=3}^N y(i-1)[y(i) + y(i-2)]}{2 \sum_{i=3}^N y(i-1)^2} \right] \quad (17)$$

where $y(i); i=1,2,3, \dots, N$ are the measured samples of the cosine wave. With $N = 50$, the method was evaluated using 500 noisy cosine waves having the same amplitude, period, and SNR as those used for AIM evaluation and CRLB determination. Although there is no sharp cut off, estimation accuracy for this method deteriorates steadily, departing more from the CRLB bound as the noise level increases. At $SNR = 0$ dB, the mean absolute percentage frequency error is unacceptably high at 77 % for this method, as compared to about 5 % for AIM. It is clear that AIM frequency estimates are far more accurate than those using the Prony's based method, particularly at high noise levels. Fig. 6(b) for the amplitude bounds shows that AIM estimation errors for the amplitude track the CRLB well at low values of the signal to noise ratios, with a loss of only about 6 dB at $SNR = 0$ dB. Again there are no signs of thresholding at high noise levels.

We compared the performance of the AIM frequency estimator and the DFT based method for determining the frequency of a single noisy sine wave having $A = 4.58$ and $T = 28$ at $SNR = 7$ dB, see Fig. 7. An FFT transform was performed on the waveform signal, and the frequency of the sine wave was determined as that corresponding to the frequency bin containing the peak magnitude of the FFT transform. Percentage errors in the period estimates for the AIM and DFT methods are 0.74% and -10.71%, respectively. The AIM amplitude estimator was compared with a simple peak detection method applied after the waveform was smoothed to reduce the effects of noise. Smoothing was performed through simple averaging of the waveform signal over a moving window of size 3. Both the positive and negative peak values were detected, and the amplitude of the sine wave was taken as the average of their absolute values. Percentage

errors in the amplitude estimates for the AIM and peak detection methods are -9.3% and 10.3% , respectively. Fig. 7 plots the original sine wave data in comparison with two sine wave fits obtained from period and amplitude estimates by the AIM method and the DFT-Peak detection method. Computed to give an overall measure of the goodness of fit, the sum of the squared errors was 32 and 289 for the two fits, respectively.

7. Conclusions

Abductive machine learning has been demonstrated for estimating the amplitude and frequency of a single noisy sinusoid. It promises faster, simpler, and more automated analysis that makes it favorable in many real-time applications. Both parameters can be accurately determined simultaneously through direct substitution of the measured sample values into a set of polynomial equations, without the need for user intervention. Parameter estimation does not require any matrix manipulations, data fitting, or solving for the roots of polynomials, and therefore can be performed speedily in the field using simple portable apparatus. Major computational resources are needed only once for training during model synthesis. Performance for both amplitude and frequency is superior to that of a back propagation neural network model developed on the same noisy data. Frequency estimation accuracy exceeds that of conventional DFT based techniques, particularly when data is available only on a small portion of the sinusoid cycle. Performance is also superior to that of a neural network model, and that of a Prony's based method using a linear least squares fit, particularly at high noise levels. AIM estimates start to show some bias at about $SNR = 3$ dB. Error variance tracks that of the Cramer-Rao bound closely (with a small loss) at low SNR values, showing no signs of thresholding down to $SNR = 0$ dB. This may be attributed to the fact that phase was assumed constant and that a model was synthesized for use at each noise level considered. Future work would consider estimating the phase for a given frequency as well as the analysis of multiple sinusoids.

Acknowledgements

This work is part of KFUPM/RI Center for Applied Physical Sciences project supported by the Research Institute of King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

References

Abdel-Aal, R.E., 1998. Automatic fitting of Gaussian peaks using adaptive machine learning. *IEEE Transactions on Nuclear Science* 45, 1-16.

Abdel-Aal, R.E., Elhadidy, M.A., 1994. A machine-learning approach to modeling and forecasting the minimum temperature at Dhahran, Saudi Arabia. *Energy- The International Journal* 19, 739-749.

AbTech Corporation, 1990. AIM User's Manual, Charlottesville, VA.

Barron, A.R., 1984. Predicted squared error: A criterion for automatic model selection. In: Farlow, S.J. (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, Marcel-Dekker, New York, 87-103.

Barron, R.L., Mucciardi, A.N., Cook, F.J., Craig, J.N, Barron, A.R., 1984. Adaptive learning networks: Development and application in the United States of algorithms related to GMDH. In: Farlow, S.J. (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, Marcel-Dekker, New York, 25-65.

Farlow, S.J., 1984. The GMDH algorithm. In: Farlow, S.J. (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, Marcel-Dekker, New York, 1-24.

Garcia, G., 2001. Estimation of sinusoids in audio signals using an analysis-by-synthesis neural network. IEEE International Conference on Acoustics, Speech, and Signal Processing, 3369-3372.

Hancke, G.P., 1990. The optimal frequency estimation of a noisy sinusoidal signal. IEEE Transactions on Instrumentation and Measurements 39, 843-846.

Hummels, D.M., Ahmed, W., Musavi, M.T., 1995. Adaptive detection of small sinusoidal signals in non Gaussian noise using an RBF neural network. IEEE Transaction on Neural networks 6, 214-219.

Karhunen, J., Joutsensalo, J., 1992. Learning of sinusoidal frequencies by nonlinear constrained Hebbian algorithms. The IEEE-SP workshop on Neural Networks for Signal Processing, 39-48.

Kay, S., 1989. A fast and accurate single frequency estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing 37, 1987-1990.

Kay, S.M., Marple Jr. S.L., 1981. Spectrum analysis- A modern perspective. Proceedings of the IEEE 69, 1380-1419.

Marple Jr., S.L., 1989. A tutorial review of modern spectral estimation. International Conference on Acoustics, Speech, and Signal Processing, 2152-2157.

Mendenhall, W., Beaver, R. J., 1994. Introduction to Probability and Statistics, Duxbury Press, Belmont, California.

Montgomery, G.J., Drake, K.C., 1990. Abductive Networks. SPIE Applications of Artificial Neural Networks Conference, 56-64.

Moreno, A., Fonollosa, J.A.R., 1992. Pitch determination of noisy speech using higher order statistics. International Conference on Acoustics, Speech, and Signal Processing, 133-136.

Nishiyama, K., 1997. A nonlinear filter for estimating a sinusoidal signal and its parameters in white noise: On the case of a single sinusoid. IEEE Transactions on Signal Processing 45, 970-981.

Nishiyama, K., 1999. Robust estimation of a single complex sinusoid in white noise-filtering approach. IEEE Transactions on Signal Processing 47, 2853-2856.

Norton, J.P., 1986. An introduction to Identification, Academic Press, London.

Osowski, S., 1992. Neural network for estimation of harmonic components in a power system. IEE Proceedings, Pt. C 139, 129-135.

Rao, S.S., Raghavan, K.R., 1989. Estimating the frequencies of sinusoids at low signal to noise ratios using the principal component method in tandem with the Burg algorithm. International Conference on Acoustics, Speech, and Signal Processing, 2210-2213.

Rife, D.C., Boorstyn, R.R., 1974. Single-tone parameter estimation from discrete time observations. IEEE Transactions on Information Theory IT-20, 591-598.

Rife, D.C., Boorstyn, R.R., 1976. Multiple tone parameter estimation from discrete time observations. *Bell System Technical Journal* 55, 1389-1410.

Sills, J.A., Black, Q.R., 1996. Frequency estimation from short pulses of sinusoidal signals. *Military Communications Conference (MILCOM)*, 979-983.

Stoica, P., Moses, R.L., Friedlander B., Soderstrom T., 1989. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 378-392.

Tufts, D.W., Fiore, P.D., 1996. Simple effective estimation of frequency based on Prony's method. *International Conference on Acoustics, Speech, and Signal Processing*, 2801-2804.

Weiss, S.M., Kulikowski, C.A., 1991. *Computer Systems That Learn*. Morgan Kaufmann, CA.

Table 1. Error statistics for the AIM models of Figure 2 for fitting a single noiseless sine wave at various frequency ranges. Evaluation on 500 noiseless cases. Shown are percentages of the evaluation population having various ranges of the absolute percentage error and data on the mean and standard deviation of the absolute percentage error.

Data for Training and Evaluation	Period Range (Seconds)	Amplitude					Period				
		% Error Population			Absolute % Error		% Error Population			Absolute % Error	
		±2%	±5%	±10%	Mean	SD	±2%	±5%	±10%	Mean	SD
1-decade of Frequency Variations	5-50	47.2	77.2	93.2	3.56	4.04	46.8	82.8	95.8	3.74	12.73
1-Octave of High Frequency Variations	5-10	100	100	100	0.72	0.42	91.8	95.4	99.8	0.77	1.11
1-Octave of Medium Frequency Variations	25-50	100	100	100	0.31	0.18	94.2	99.1	100	0.79	0.90
1-Octave of Low Frequency Variations	100-200	72.2	98	100	1.59	1.33	86.5	98	100	1.13	1.20

Table 2. Error statistics for the AIM models of Figure 4 for fitting a single sine wave at various noise levels. Evaluation on 500 noisy cases. Shown are percentages of the evaluation population having various ranges of the absolute percentage error and data on the mean and standard deviation of the absolute percentage error.

Data for Training and Evaluation	Amplitude					Period				
	% Error Population			Absolute % Error		% Error Population			Absolute % Error	
	±2%	±5%	±10%	Mean	SD	±2%	±5%	±10%	Mean	SD
Noiseless Data	100	100	100	0.31	0.18	94.2	99.1	100	0.79	0.90
Noisy Data, SNR = 17 dBs	40.7	81.5	98.2	3.04	2.28	72.8	96.7	99.6	1.70	3.07
Noisy Data, SNR = 3 dBs	11.7	27.5	49.1	12.13	9.62	30.7	63.7	91.6	4.53	3.88

Table 3. Data for testing the hypothesis of unbiased estimation for the AIM amplitude and period estimators using the z statistic at two values of the signal to noise ratio. Mean and standard deviation are those of the percentage error for a 500-case evaluation sample.

SNR, dBs	Amplitude			Period		
	Sample Mean, \bar{x}	Sample Standard Deviation, SD	z Statistic, $z = \frac{\bar{x} - 0}{SD/\sqrt{n}}$	Sample Mean, \bar{x}	Sample Standard Deviation, SD	z Statistic, $z = \frac{\bar{x} - 0}{SD/\sqrt{n}}$
7	0.3777	10.5417	0.801	0.2167	4.7676	1.016
3	2.0512	15.3587	2.986	0.6116	5.9378	2.303

Table 4. Performance of AIM models when evaluated with data having the same and different noise levels from those used to develop the model. Shown are the mean and standard deviation of the absolute percentage error for both the amplitude and the period.

Evaluation Data Training Data Data	Noiseless Data				Noisy Data, SNR = 17 dBs				Noisy Data, SNR = 7 dBs			
	Amplitude		Period		Amplitude		Period		Amplitude		Period	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Noiseless Data	0.31	0.18	0.79	0.90	2.95	2.24	4.22	5.14	28.60	20.94	715.06	5144.14
Noisy Data, SNR = 17 dBs	0.75	0.53	0.97	1.10	3.04	2.28	1.70	3.07	19.80	15.49	31.83	157.91
Noisy Data, SNR = 7 dBs	2.40	2.23	1.78	1.76	2.62	2.26	1.70	1.97	8.41	6.36	3.35	3.40

Table 5. Performance comparison between the AIM abductive model shown in the middle row of Figure 4 for a single noisy sine wave at $SNR = 17$ dB and a back propagation neural network model developed and evaluated using the same data.

Model	Amplitude					Period				
	% Error Population			Absolute % Error		% Error Population			Absolute % Error	
	$\pm 2\%$	$\pm 5\%$	$\pm 10\%$	Mean	SD	$\pm 2\%$	$\pm 5\%$	$\pm 10\%$	Mean	SD
Abductive Model (Middle row in Fig. 4)	40.7	81.5	98.2	3.04	2.28	72.8	96.7	99.6	1.70	3.07
Neural Network Model (50-3-2)	28.6	65.8	92.4	4.40	4.00	34.2	87.8	98.8	2.53	2.52

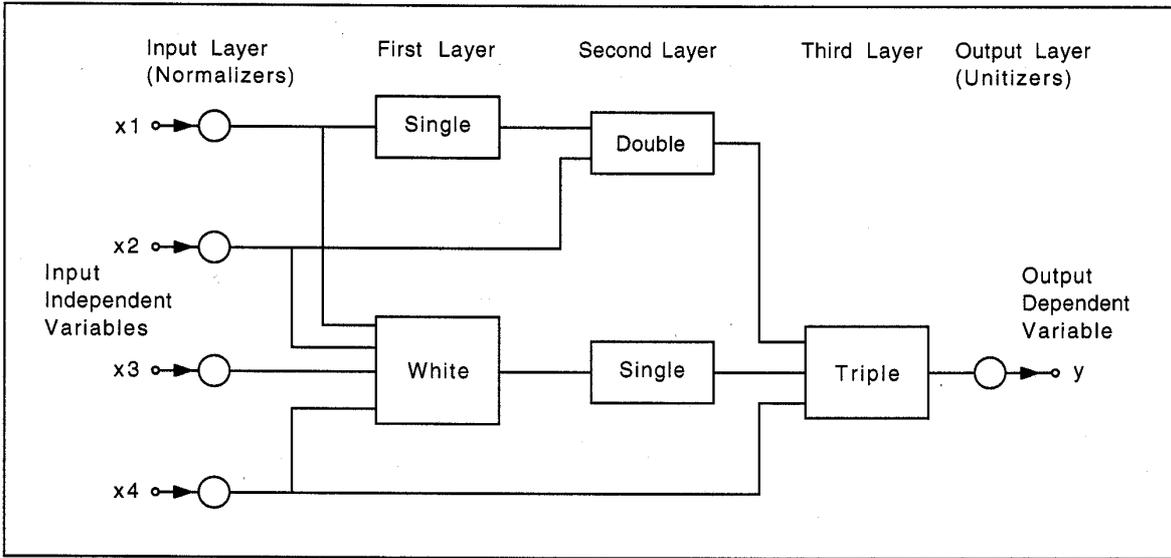


Figure 1. A typical AIM abductive network structure showing various types of functional elements.

Training Data	Amplitude	Period
1-decade of Frequency Variations Period Range: 5-50		
1-Octave of High Frequency Variations Period Range: 5-10		
1-Octave of Medium Frequency Variations Period Range: 25-50		
1-Octave of low Frequency Variations Period Range: 100-200		

Figure 2. AIM abductive network models for estimating the parameters of a single noiseless sine wave at various frequency ranges. Amplitude range: 1-10. Training on 1000 cases.

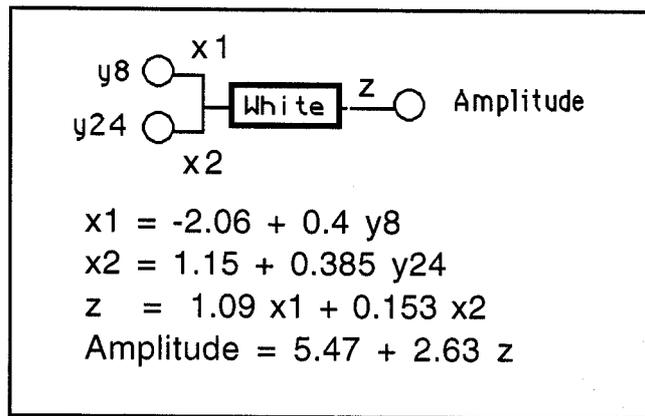


Figure 3. Details of the AIM abductive network model for the amplitude of a single noiseless sine wave (third row, left, in Figure 2).

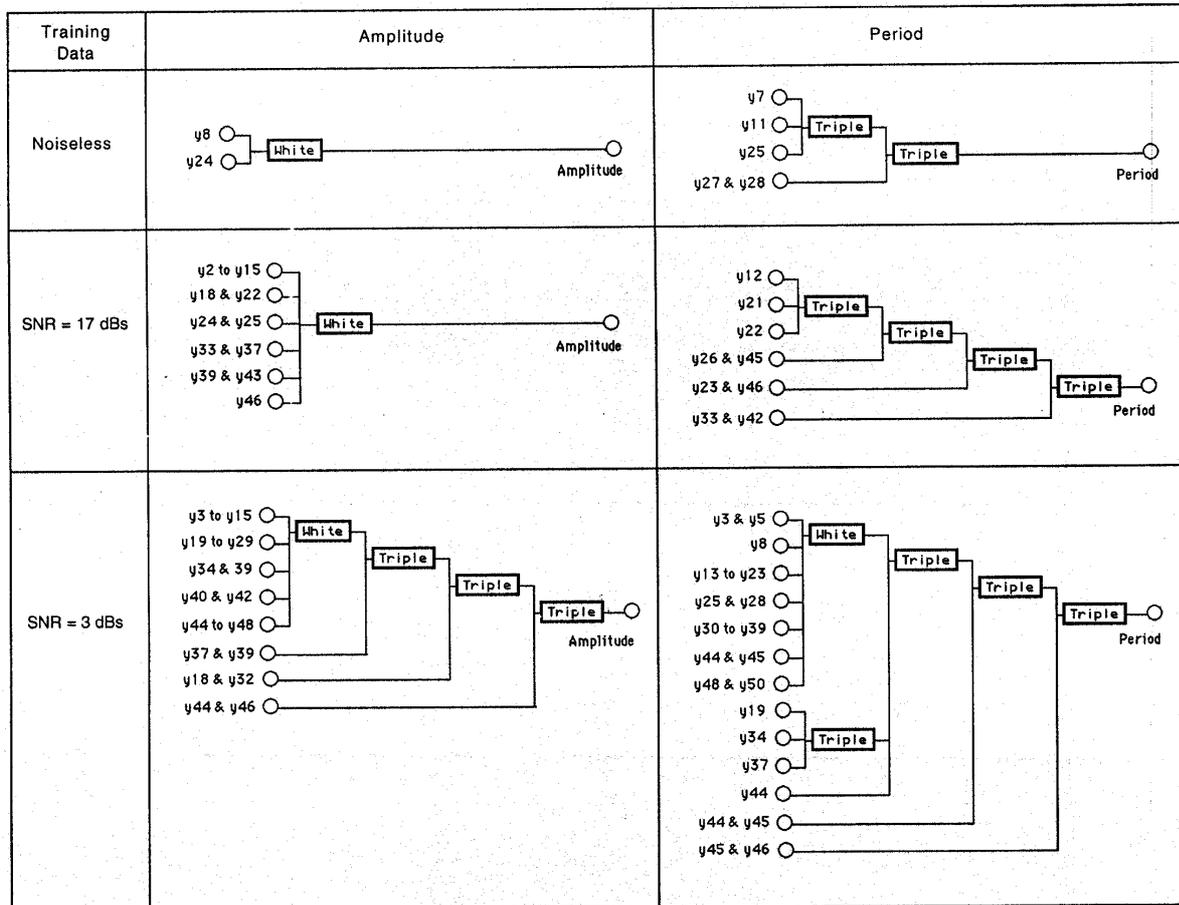


Figure 4. AIM abductive network models for estimating the parameters of a single sine wave at $SNR = 17$ dB and $SNR = 3$ dB in comparison with the noiseless case. Training data: Medium frequency octave variations, 1000 cases.

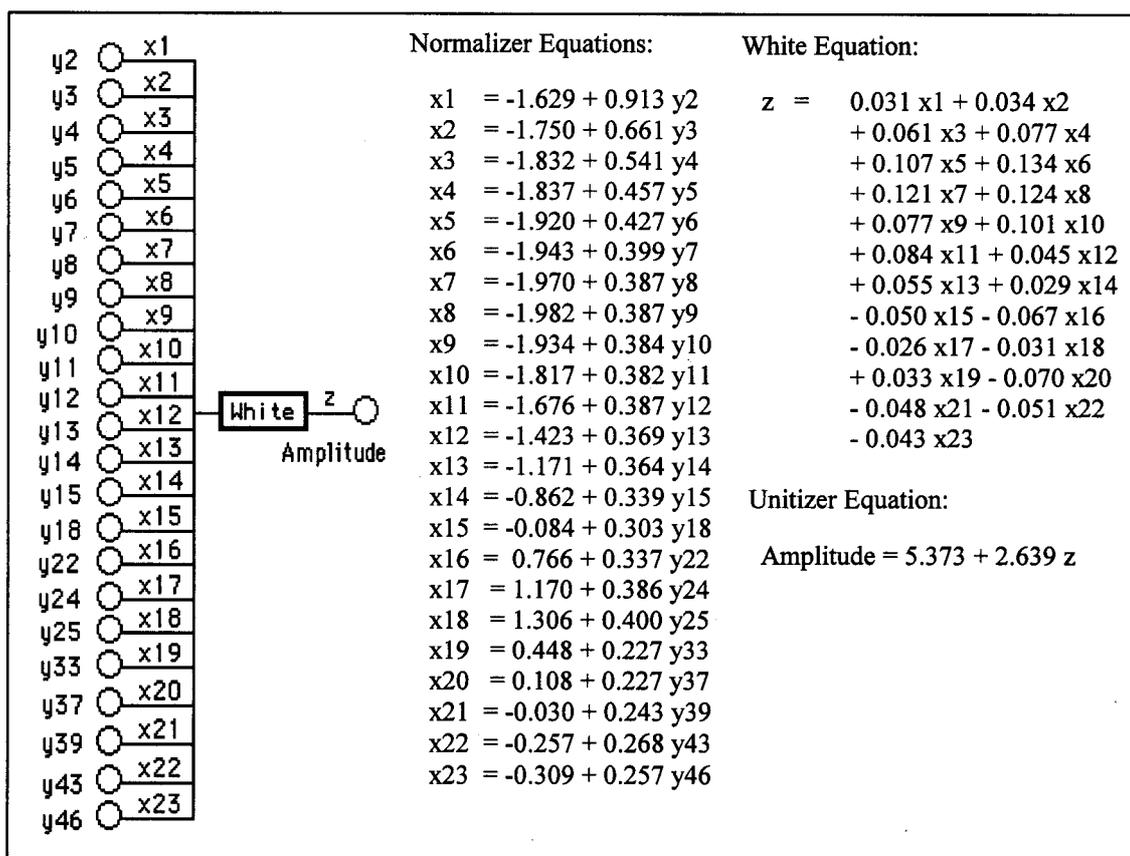
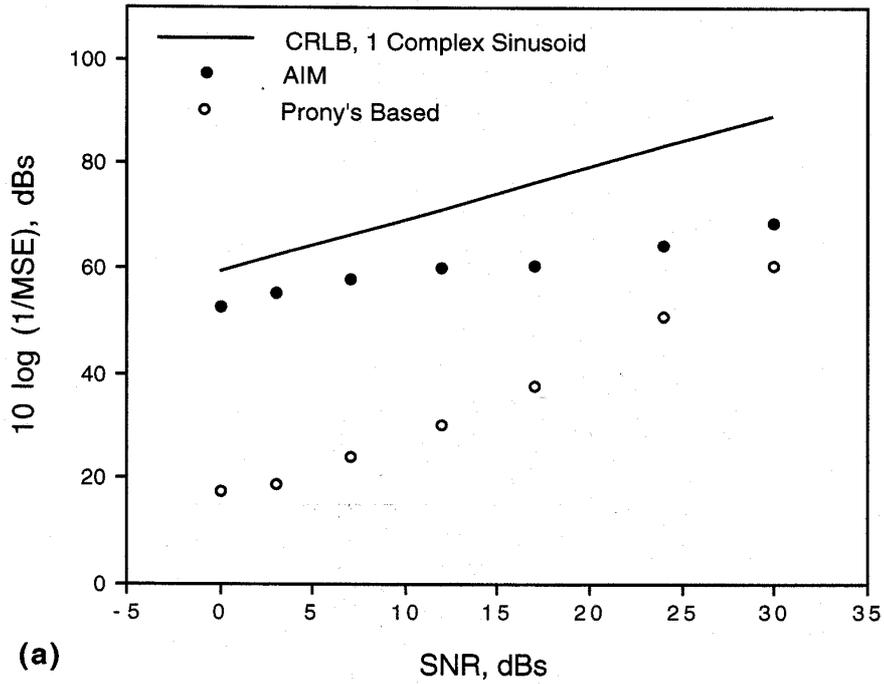
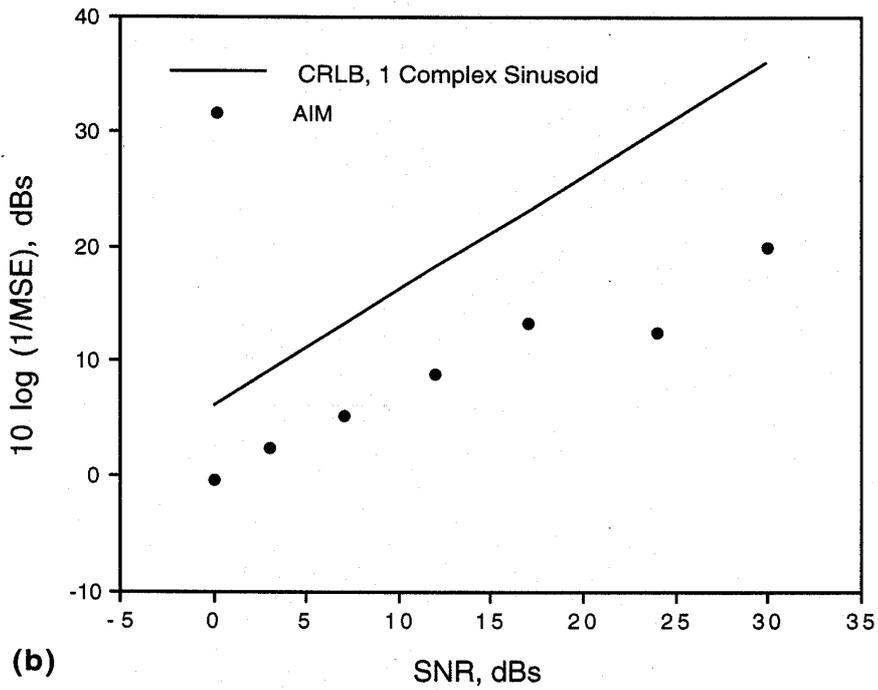


Figure 5. Details of the AIM abductive network model for the amplitude of a single noisy sine wave at SNR = 17 dB (middle row, left, in Figure 4).



(a)



(b)

Figure 6. Plots of the inverse of estimation error variance versus signal-to-noise ratio for the Cramer-Rao bounds, AIM estimators, and a Prony's based frequency estimator. (a): Frequency, (b): Amplitude. Evaluation on 500 cases having amplitude = 5 and period = 26 at the same SNR used to develop the model.

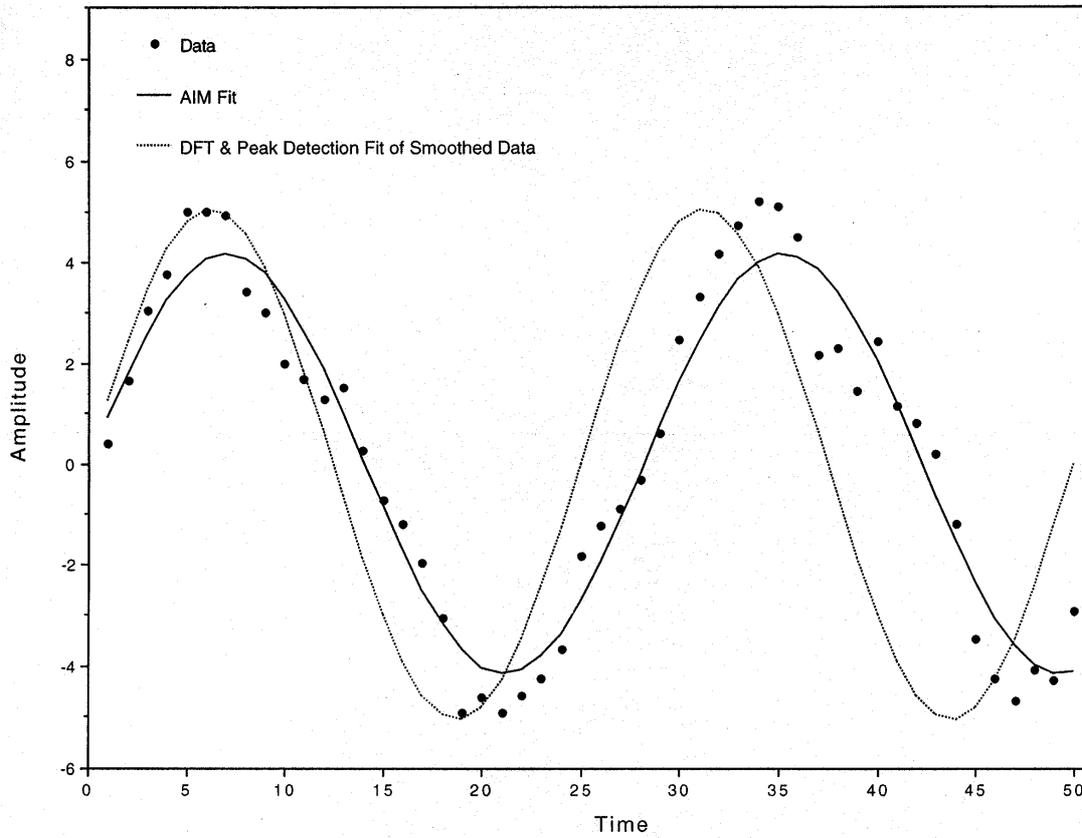


Figure 7. Comparison between fits for a noisy sine wave ($A = 4.58$, $T = 28$, $SNR = 7$ dB) using AIM amplitude and period estimators and using DFT and peak detection to determine the frequency and the amplitude, respectively.