Bayesian Autoregressive Modelling for Short Sequences

Douglas E. Johnston and Petar M. Djurić

CEAS Technical Report #669

Department of Electrical Engineering

State University of New York at Stony Brook

Stony Brook, NY 11794, USA

August 11, 1993

Abstract

A Bayesian solution to autoregressive (AR) signal modelling is proposed. In contrast to standard techniques, the *exact* posterior density of the model parameters is derived. In addition, a numerically efficient procedure is introduced which permits application of this density to compute optimal estimates, such as minimum mean square estimates (MMSE), without resorting to multi-dimensional optimization. Our approach can be used to investigate many signal characteristics such as the signal's spectrum, marginal densities for prediction or model selection. The end result is not only guaranteed to be more accurate, but also more robust to model assumptions. Simulation results confirm our expectations and illustrate the improvement over the classic, maximum conditional likelihood (MCL) approach to AR signal modelling.

1 Introduction

The autoregressive (AR) signal model has long been utilized by the signal processing community [11] and, even today, the AR model continues to be an active area of research, particularly for small sample sizes [4, 5]. This is an intriguing problem, not only due to its inherent difficulty, but also due to its widespread applicability. In many applications, such as speech processing, ECG analysis, or econometrics, the signals are highly nonstationary which implies that short data segments are the norm for signal analysis. In other applications, such as geophysics or sonar, the limiting factor is the number of sensors that observe the phenomena. A particular example of where the AR model has been successfully applied is the modelling of EEG signals for biomedical applications [1].

Generally speaking, there have been two popular approaches to AR signal modelling. The first approach is based on the fact that the signal covariances satisfy the Yule-Walker equations [13]. The method proceeds by computing an estimate of the signal covariances, based on the observed data, and then solving the resulting matrix equation. This approach has the advantage of having very efficient algorithms for solving for the AR coefficients and works very well as long as the covariance estimates are reliable, which typically requires a substantial number of data points.

The other popular approach to solving this problem is based on classical statistics or, equivalently, the likelihood function. This method has the benefit of utilizing the model of the driving noise process though, unfortunately, the exact likelihood is a complicated function of the AR parameters which is intractable to maximize, analytically. To circumvent this difficulty, many researchers use an approximate likelihood [10] which is conditioned on the first p samples, where p is the model order. By using this approximation, a simple solution exists for the maximum likelihood estimate, which we denote as MCL to differentiate it from the exact ML estimate. The disadvantage is, the approximation is an asymptotic one which implies that it will work well for large data records though all bets are off when the number of samples

is small.

Recently, Bayesian methods have been applied to the AR signal model [3, 16], though the conditional likelihood is utilized since it results in a rather elegant solution. The nicest property is that, if noninformative priors are used and the signal variance is known, the posterior density is Gaussian and hence convenient to work with. In addition, Kalman filter methods can be applied which makes this approach amenable to real-time processing [8]. Once again, the dilemma is that this solution is only guaranteed to give reasonable results when sufficient data are available.

In our approach, we too will utilize a Bayesian methodology though, in contrast, we will construct the exact posterior density of the AR model parameters. By deriving an expression for the theoretical signal covariance matrix in terms of the AR coefficients, we are able to write the true likelihood function directly. In addition, since the variance of the driving noise process is unknown, it is marginalized analytically using a noninformative prior.

By avoiding asymptotic results, we derive a procedure to utilize the posterior density to examine various signal characteristics, such as its spectrum. This is in contrast to the traditional methods, which first estimate the parameters of the model and then use those estimates in some desired fashion, such as prediction or control [2]. To implement the required integrals, we introduce a numerically efficient algorithm, importance sampling, which not only guarantees optimal estimates but also avoids the drawbacks associated with multidimensional optimization. Simulation results illustrate our proposed method and the improvement that can be attained compared with the MCL approach, particularly for short data records. The Bayesian method, coupled with numerical integration techniques, offers serious contention to classical approaches, not only for AR signal modelling, but for many other estimation problems as well.

The paper is organized as follows. In Section 2, the problem is formulated and the necessary assumptions are stated. Following that, in Section 3, we present our

derivations and construct the posterior distribution of the AR model parameters. The importance sampling method is presented in Section 4 as well as to how to best take advantage of its benefits. In Section 5, simulation results are reported which clearly illustrate the benefit of our proposed approach and finally, in Section 6, we present our conclusions, and suggestions for future research.

2 Problem Formulation

A real signal, $y(\cdot)$, is observed from N data samples. The signal is assumed to be generated as a stationary, p'th order AR process, where p may be unknown. We may thus model the signal as

$$y(n) = \sum_{j=1}^{p} a_j y(n-j) + e(n) \qquad n = 1, 2, \cdots, N \tag{1}$$

where $e(n)$ is a sample from a white Gaussian process, with zero mean and unknown variance σ^2, and $a_j, j = 1, \cdots, p$, are the coefficients of the process.

It is well known [13], that the Z-transform of the signal is given by

$$H(z) = \frac{\sigma^2}{|1 - \sum_{j=1}^{p} a_j z^{-j}|^2}, \tag{2}$$

where z is a complex variable, and that the poles of the process correspond to the zeroes of the characteristic polynomial. That is, they are the solutions to the equation

$$z^p - a_1 z^{p-1} - a_2 z^{p-2} \cdots - a_p = 0. \tag{3}$$

The stationary assumption of the model implies the constraint that the poles of the process reside within the unit circle in the complex plane. This, in turn, constrains the coefficients, $a_j, j = 1, \cdots, p$, to lie in a region $\Omega \subset \Re^p$, which we will refer to as the region of stationarity.

Our objective, given the model, is to determine the exact posterior density of the AR coefficients given the observed data, which is written as $f(\mathbf{a}|\mathbf{y}_{1,N})$ and we have

used the notation

$$\mathbf{a} = [a_1, a_2, \cdots, a_p]^T \tag{4}$$

and

$$\mathbf{y}_{j,k} = [y(j), y(j+1), \cdots, y(k)]^T \qquad j \leq k. \tag{5}$$

Upon deriving this density, information about the signal is to be extracted by integrating the desired function using the density as a weighting function. For example, the MMSE spectral estimate is computed as

$$\hat{S}(w) = \int_{\mathbf{a}} S(w|\mathbf{a}) f(\mathbf{a}|\mathbf{y}_{1,N}) d\mathbf{a} \tag{6}$$

where $S(w|\mathbf{a})$ is the spectrum of an AR process whose coefficients are given by $\mathbf{a}$. Another example is the one-step predictive density given as

$$f(y(N+1)|\mathbf{y}_{1,N}) = \int_{\mathbf{a}} f(y(N+1)|\mathbf{a}, \mathbf{y}_{1,N}) f(\mathbf{a}|\mathbf{y}_{1,N}) d\mathbf{a} \tag{7}$$

which can be used for decision making, incorporating appropriate cost functions. It should be noted that our approach is significantly different than the traditional method of first estimating the coefficients and then using them to interrogate the signal. Using the true posterior density, and marginalizing over all possible values of $\mathbf{a}$, not only guarantees improved performance (i.e., minimum mean square estimates) but also robustness to excursions from the assumed model.

3 Theoretical Derivation

To proceed, we will use Bayes' rule to evaluate the posterior density. That is,

$$f(\mathbf{a}|\mathbf{y}_{1,N}) = \frac{f(\mathbf{y}_{1,N}|\mathbf{a}) f(\mathbf{a})}{\int_{\mathbf{a}} f(\mathbf{y}_{1,N}|\mathbf{a}) f(\mathbf{a}) d\mathbf{a}} \tag{8}$$

where $f(\mathbf{a})$ is the prior density for the AR coefficients. Note, that since the variance, σ^2, is unknown, it must be integrated out of the likelihood function as a nuisance parameter. That is,

$$f(\mathbf{y}_{1,N}|\mathbf{a}) = \int_{\sigma} f(\mathbf{y}_{1,N}|\mathbf{a}, \sigma) f(\sigma|\mathbf{a}) d\sigma \tag{9}$$

and, therefore, we can write (8) as

$$f(\mathbf{a}|\mathbf{y}_{1,N}) = \frac{\int_\sigma f(\mathbf{y}_{1,N}|\mathbf{a},\sigma)f(\sigma|\mathbf{a})d\sigma f(\mathbf{a})}{\int_\mathbf{a}\int_\sigma f(\mathbf{y}_{1,N}|\mathbf{a},\sigma)f(\sigma|\mathbf{a})f(\mathbf{a})d\sigma d\mathbf{a}} \tag{10}$$

We would like to justly compare our approach with other techniques and thus we will adopt non-informative priors by making use of Jeffreys' invariance rule [3]. This implies that $f(\mathbf{a})$ is a constant over the region of stationarity, Ω, and that $f(\sigma|\mathbf{a}) \propto \frac{1}{\sigma}$. Other Bayesian approaches to AR signal modelling [5, 16] utilize the improper prior, $f(\mathbf{a}) = C$, a constant, over $\Re^p$ which, of course, does not imply the signal is stationary. Also, standard maximum likelihood estimates of $\mathbf{a}$ can be thought of as maximum a posteriori (MAP) estimates when the improper prior is used.

The likelihood function, $f(\mathbf{y}_{1,N}|\mathbf{a},\sigma)$, needs to be determined and, to facilitate this, we can decompose it as

$$f(\mathbf{y}_{1,N}|\mathbf{a},\sigma) = f(\mathbf{y}_{p+1,N}|\mathbf{y}_{1,p},\mathbf{a},\sigma)f(\mathbf{y}_{1,p}|\mathbf{a},\sigma). \tag{11}$$

The first term, in (11), is the conditional likelihood and is written as

$$f(\mathbf{y}_{p+1,N}|\mathbf{y}_{1,p},\mathbf{a},\sigma) = \frac{1}{(2\pi\sigma^2)^{(N-p)/2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}_{p+1,N} - \mathbf{Y}_{p,N}\mathbf{a})^T(\mathbf{y}_{p+1,N} - \mathbf{Y}_{p,N}\mathbf{a})\right\} \tag{12}$$

where we have introduced the $(N-p) \times p$ matrix,

$$\mathbf{Y}_{p,N} = \left[\mathbf{y}_{p,N-1}\vdots\cdots\vdots\mathbf{y}_{1,N-p-1}\right]. \tag{13}$$

In the traditional approach to AR signal modelling, an asymptotic argument is utilized and the conditional likelihood is used to construct point estimates of the AR coefficients. The main benefit to this approach is that the conditional likelihood is in a form which is equivalent to when a linear data model is assumed and, therefore, the MCL estimate has a simple and well known solution [9]. In addition, if σ is assumed known and a flat prior is used for $\mathbf{a}$, the posterior density can be shown to be Gaussian [3] which is ideal for computing point estimates and confidence ellipsoids. The drawback is that this approximation is asymptotic and, thus, only valid for large data records.

The second term in (11) is the likelihood of the first p samples, the evaluation of which entails expressing the theoretical covariances in terms of the AR coefficients. This, of course, complicates the exact likelihood which is why it is traditionally ignored. Not only does the model not conform to a linear hypothesis, but also the posterior is non-Gaussian even when σ is known. Let $\mathbf{R}_p$ denote the $p \times p$ covariance matrix,

$$\mathbf{R}_p = \begin{bmatrix} R(0), & R(1), & \cdots, & R(p-1) \\ R(1), & R(0), & \cdots, & R(p-2) \\ \vdots, & \vdots, & \ddots & \vdots \\ R(p-1), & R(p-2), & \cdots, & R(0) \end{bmatrix}, \tag{14}$$

whose i,j'th element is $R(i-j) = E[y(i)y(j)]$. It should be noted that $\mathbf{R}_p$ is Toeplitz and therefore many fast, numerical algorithms are available for calculating its inverse and determinant.

The Yule-Walker equation [2] for the signal covariances is written as

$$\mathbf{R}_{p+1}\tilde{\mathbf{a}} = \sigma^2 \mathbf{e}_1 \tag{15}$$

where $\tilde{\mathbf{a}}^T = [1, \mathbf{a}^T]$ and $\mathbf{e}_1$ is the standard, unit coordinate vector. The matrix $\mathbf{R}_{p+1}$ can be written as

$$\mathbf{R}_{p+1} = \begin{bmatrix} \mathbf{r}_{p+1}^T \mathbf{H}_0 \\ \mathbf{r}_{p+1}^T \mathbf{H}_1 \\ \vdots \\ \mathbf{r}_{p+1}^T \mathbf{H}_{p+1} \end{bmatrix} \tag{16}$$

where $\mathbf{r}_{p+1}^T = [R(0), R(1), \cdots, R(p)]$ and the $\mathbf{H}_k$'s are constructed, recursively, via simple linear operations. Letting $\mathbf{I}_q$ denote the $q \times q$, identity matrix and $\mathbf{0}_{m,n}$ the $m \times n$, zero matrix, the $\mathbf{H}_k$'s can be written as $\mathbf{H}_0 = \mathbf{I}_{p+1}$ and

$$\mathbf{H}_k = \mathbf{H}_{k-1}\mathbf{I}_1 + \mathbf{I}_2\mathbf{H}_{k-1}\mathbf{I}_3 \tag{17}$$

for $k = 1, \cdots, p+1$, where

$$\mathbf{I}_1 = \begin{bmatrix} \mathbf{0}_{p,1} & \mathbf{I}_p \\ 0 & \mathbf{0}_{1,p} \end{bmatrix}, \tag{18}$$

$$\mathbf{I}_2 = \begin{bmatrix} \mathbf{0}_{1,p} & 0 \\ \mathbf{I}_p & \mathbf{0}_{p,1} \end{bmatrix}, \tag{19}$$

and

$$\mathbf{I}_3 = \begin{bmatrix} 1 & \mathbf{0}_{1,p} \\ \mathbf{0}_{p,1} & \mathbf{0}_{p,p} \end{bmatrix}. \tag{20}$$

Thus, $\mathbf{H}_k$ is constructed by shifting all the columns of $\mathbf{H}_{k-1}$ to the right and replacing the first column with that of $\mathbf{H}_{k-1}$, shifted down one.

Using the above construction, we may write the Yule-Walker equation as

$$\begin{bmatrix} \mathbf{r}_{p+1}^T \mathbf{H}_0 \tilde{\mathbf{a}} \\ \mathbf{r}_{p+1}^T \mathbf{H}_1 \tilde{\mathbf{a}} \\ \vdots \\ \mathbf{r}_{p+1}^T \mathbf{H}_{p+1} \tilde{\mathbf{a}} \end{bmatrix} = \sigma^2 \mathbf{e}_1 \tag{21}$$

or, by rearranging terms,

$$\begin{bmatrix} \tilde{\mathbf{a}}^T \mathbf{H}_0^T \\ \tilde{\mathbf{a}}^T \mathbf{H}_1^T \\ \vdots \\ \tilde{\mathbf{a}}^T \mathbf{H}_{p+1}^T \end{bmatrix} \mathbf{r}_{p+1} = \sigma^2 \mathbf{e}_1 \tag{22}$$

and, thus, $\mathbf{r}_{p+1}$ can be solved for as

$$\mathbf{r}_{p+1} = \sigma^2 \mathbf{A}^{-1} \mathbf{e}_1 \tag{23}$$

where

$$\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{a}}^T \mathbf{H}_0^T \\ \tilde{\mathbf{a}}^T \mathbf{H}_1^T \\ \vdots \\ \tilde{\mathbf{a}}^T \mathbf{H}_{p+1}^T \end{bmatrix}. \tag{24}$$

If, in addition, we introduce the matrix $\tilde{\mathbf{I}} = [\mathbf{I}_p, 0]$, and noting that $\mathbf{r}_p = \tilde{\mathbf{I}} \mathbf{r}_{p+1}$ then

the covariance matrix, $\mathbf{R}_p$, is written as

$$\mathbf{R}_p = \sigma^2 \begin{bmatrix} (\check{\mathbf{I}}\mathbf{A}^{-1}\mathbf{e}_1)^T\mathbf{H}_0 \\ (\check{\mathbf{I}}\mathbf{A}^{-1}\mathbf{e}_1)^T\mathbf{H}_1 \\ \vdots \\ (\check{\mathbf{I}}\mathbf{A}^{-1}\mathbf{e}_1)^T\mathbf{H}_p \end{bmatrix}. \tag{25}$$

Hence, we may express the likelihood of the first p samples as

$$f(\mathbf{y}_{1,p}|\mathbf{a},\sigma) = \frac{1}{(2\pi\sigma^2)^{p/2}|\mathbf{R}_{p,\mathbf{a}}|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}_{1,p}^T\mathbf{R}_{p,\mathbf{a}}^{-1}\mathbf{y}_{1,p}\right\}. \tag{26}$$

where the matrix $\mathbf{R}_{p,\mathbf{a}} = \frac{1}{\sigma^2}\mathbf{R}_p$ is the correlation coefficient matrix, defined via (25), and we have explicitly noted its dependence on $\mathbf{a}$. Given a value of $\mathbf{a}$, $\mathbf{R}_{p,\mathbf{a}}$ is easily determined via (25) and its inverse and determinant can be efficiently computed. As previously noted, most research to date has ignored this term, which is reasonable for large sample sizes. For short data sequences, however, this term is significant and can not be ignored.

Combining (12) and (26), the complete likelihood of the AR coefficients can be written as

$$\begin{aligned} f(\mathbf{y}_{1,N}|\mathbf{a},\sigma) &= \tfrac{1}{(2\pi\sigma^2)^{N/2}|\mathbf{R}_{p,\mathbf{a}}|^{1/2}} \\ \times \exp\left\{-\tfrac{1}{2\sigma^2}((\mathbf{y}_{p+1,N} - \mathbf{Y}_{p,N}\mathbf{a})^T(\mathbf{y}_{p+1,N} - \mathbf{Y}_{p,N}\mathbf{a}) + \mathbf{y}_{1,p}^T\mathbf{R}_{p,\mathbf{a}}^{-1}\mathbf{y}_{1,p})\right\}. \end{aligned} \tag{27}$$

The marginalization of σ can be carried out by using the integral formula

$$\int_0^\infty \frac{1}{\sigma^{2L}} e^{-\frac{\beta}{\sigma^2}} \frac{d\sigma}{\sigma} = \frac{\Gamma(L)}{2\beta^L} \tag{28}$$

which enables us to write

$$f(\mathbf{a}|\mathbf{y}_{1,N}) \propto \frac{|\mathbf{R}_{p,\mathbf{a}}^{-1}|^{1/2}}{\left((\mathbf{y}_{p+1,N} - \mathbf{Y}_{p,N}\mathbf{a})^T(\mathbf{y}_{p+1,N} - \mathbf{Y}_{p,N}\mathbf{a}) + \mathbf{y}_{1,p}^T\mathbf{R}_{p,\mathbf{a}}^{-1}\mathbf{y}_{1,p}\right)^{N/2}} \tag{29}$$

where $\propto$ is the proportionality sign. We may write the posterior density in a more convenient form by introducing

$$\hat{\mathbf{a}} = (\mathbf{Y}_{p,N}^T\mathbf{Y}_{p,N})^{-1}\mathbf{Y}_{p,N}^T\mathbf{y}_{p+1,N}, \tag{30}$$

which is the MCL estimate for the AR coefficients, and

$$\mathbf{P}^{\perp}_{\mathbf{Y}_{p,N}} = \mathbf{I}_{N-p} - \mathbf{Y}_{p,N}(\mathbf{Y}^T_{p,N}\mathbf{Y}_{p,N})^{-1}\mathbf{Y}^T_{p,N}, \tag{31}$$

which is the orthogonal projection operator that projects vectors onto the subspace orthogonal to the column space of $\mathbf{Y}_{p,N}$. Using the above notation, the posterior density of the AR coefficients is written as

$$f(\mathbf{a}|\mathbf{y}_{1,N}) \propto \frac{|\mathbf{R}^{-1}_{p,\mathbf{a}}|^{1/2}}{\left((\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{Y}^T_{p,N}\mathbf{Y}_{p,N}(\mathbf{a}-\hat{\mathbf{a}}) + \mathbf{y}^T_{1,p}\mathbf{R}^{-1}_{p,\mathbf{a}}\mathbf{y}_{1,p} + \mathbf{y}^T_{p+1,N}\mathbf{P}^{\perp}_{\mathbf{Y}_{p,N}}\mathbf{y}_{p+1,N}\right)^{N/2}}. \tag{32}$$

4 Numerical Solution

Given the posterior density, we may compute the MMSE estimate of any function of the AR model parameters, $h(\mathbf{a})$. One such possibility is $h(\mathbf{a}) = \mathbf{a}$ which would provide us with the mean of the posterior density. To compute the estimate, we merely integrate the desired function using the posterior density as a weighting function,

$$\hat{h}_{MMSE} = \int_{\mathbf{a}} h(\mathbf{a}) f(\mathbf{a}|\mathbf{y}_{1,N}) d\mathbf{a}. \tag{33}$$

In general, the required integrations will not have a closed form solution and thus, an efficient numerical procedure is needed. One approach is to use a numerical quadrature procedure [14, 15] which estimates the integral by using a grid, which is typically non-uniform, over the region of integration and local spline, or polynomial approximations which can be integrated analytically. Though highly efficient for one dimensional integration, this approach suffers from the following drawbacks:

1. Computations increase exponentially with the dimension of the integration region.

2. Error estimates are not easily computed.

3. If a more accurate estimate is required, the previous estimate is thrown out and the complete procedure must be repeated.

4. The estimate is susceptible to the spline, or polynomial, assumption.

An alternate approach is to use a Monte Carlo procedure [7] which based on the rather simple principle that a deterministic integral can be estimated via a stochastic process. This approach has been successfully applied to many problems in statistical physics and econometrics. Formally, if the expected value of a function, $h(x)$, with respect to a probability density function (pdf), $f(x)$, is to be computed, one may estimate this integral by the following,

$$H_N = \frac{1}{N}\sum_{i=1}^{N} h(x_i) \tag{34}$$

where the x_i's are samples drawn from a population governed by the pdf, $f(x)$. It can be shown that the Monte Carlo estimate is unbiased and strongly consistent. In addition, due to the Central Limit Theorem, the estimate is asymptotically Gaussian and therefore, error estimates can be easily computed[6].

In many cases, the pdf $f(x)$ is not easily sampled. In such cases we can make use of a technique known as importance sampling which uses the fact that the integral can be rewritten as

$$\int_{\mathbf{X}} h(x)f(x)dx = \int_{\mathbf{X}} \left(\frac{h(x)f(x)}{\tilde{f}(x)}\right)\tilde{f}(x)dx \tag{35}$$

where we have tacitly assumed that the integrand on the right side of (35) is $< \infty$ except possibly on a countable set of points. In addition, we require that the support of $\tilde{f}$ contains the support of f. The integral can thus be estimated via

$$\tilde{H}_N = \frac{1}{N}\sum_{i=1}^{N} h(x_i)\frac{f(x_i)}{\tilde{f}(x_i)} \tag{36}$$

where the x_i's are now drawn from the density $\tilde{f}(x)$ which is, hopefully, more readily sampled.

The importance sampling estimate enjoys all the properties of the basic Monte Carlo method, though the rate of convergence depends on the choice of the sampling function, $\tilde{f}(x)$. In general, the function should be chosen to be as "similar" to $f(x)$ as possible and, preferably, have heavier tails. In many cases, determining the appropriate $\tilde{f}(x)$ is a difficult task and, thus, adaptive techniques are a current area of research [12].

The benefits of the Monte Carlo methods, including importance sampling, are well documented [6]. Most notably, the convergence is independent of the dimension of the integration region which, as previously stated, can not be said of numerical quadrature. In fact, there is always a dimension, d, for which the Monte Carlo converges faster than any fixed quadrature rule. In addition, an error estimate is readily available, more accurate estimates require only additional sample points, and, finally, no assumption is made on the function to be integrated. In fact, one of the most successful applications of importance sampling is when the function has a singularity within the region of integration [7].

For our application, we would like to use a Gaussian distribution, with mean μ and covariance C, for $\tilde{f}$ which is simple to implement on a computer. As noted previously, other Bayesian approaches to this problem result in a Gaussian posterior when σ is known with mean equal to $\hat{a}$ and covariance $\sigma^2(\mathbf{Y}_{p,N}^T\mathbf{Y}_{p,N})^{-1}$. If we ignore the effects of $\mathbf{R}_{p,\mathbf{a}}$ in (32), we can see that the true posterior should have a maximum close to the $\hat{a}$ and its contour ellipsoids will be basically oriented in the same manner as the, above mentioned, Gaussian density. In addition, the true posterior will tend to be strongly peaked, due to the exponentiation, and therefore the Gaussian will have larger tails which is desirable. Unfortunately, we do not know the value of σ so we will use the conditional estimate

$$\hat{\sigma}^2 = \frac{1}{N-p}\mathbf{y}_{p+1,N}^T\mathbf{P}_{\mathbf{Y}_{p,N}}^{\perp}\mathbf{y}_{p+1,N}. \tag{37}$$

This is not a serious drawback since, as indicated before, we only need a sampling function which approximates the original. Thus, for $\tilde{f}(x)$, we will use a Gaussian

density with $\mu = \hat{a}$ and

$$C = \frac{1}{N-p} \mathbf{y}_{p+1,N}^T \mathbf{P}_{\mathbf{Y}_{p,N}}^{\perp} \mathbf{y}_{p+1,N} (\mathbf{Y}_{p,N}^T \mathbf{Y}_{p,N})^{-1} \tag{38}$$

It should be noted that by using the Bayesian methodology, along with the importance sampling technique, we have replaced the traditional approach of maximizing the likelihood, or posterior, with the process of integration. Multidimensional optimization procedures suffer from the same curse of dimensionality as numerical quadrature techniques. In addition, convergence problems and local extrema hinder those methods, except for some idealized cases (i.e., convex cost functions defined on compact sets). One can see that (32) is a complicated, non-linear function of the AR coefficients and that problems such as local extrema would be encountered. Not only does the Monte Carlo method guarantee convergence but also is amenable to parallel processing.

5 Simulation Results

To illustrate the improvement that can be obtained by using the Bayesian approach, a number of simulations were performed. A second order AR process was simulated for various values of a_1 and a_2 with $\sigma = 1$. If we express the poles of the process in polar form, $\rho e^{\pm j\theta}$, it can be readily shown that

$$\rho = \sqrt{-a_2} \tag{39}$$

and

$$\theta = \tan^{-1}\left[\sqrt{-(1 + 4a_2/a_1^2)}\right]. \tag{40}$$

The simulations were run for values of ρ ranging from .2 to .8 and θ from 0 to $\pi/2$ radians. Initially, the number of samples that were used was 10 which is regarded as a short data sequence. Estimates for the AR coefficients were computed using both the traditional maximum conditional likelihood (MCL) approach as well as the proposed method (MMSE).

To analyze performance, the RMS error for the AR coefficient estimates were computed and to insure statistical significance, 500 trials were performed for each run. The number of sample points used in the Monte Carlo integration was 5000. The RMS errors for three representative values of ρ and θ are shown on Tables 1-3.

	MCL	MMSE
ϵ_{a_1}	.458	.255
ϵ_{a_2}	.413	.120

Table 1: RMS errors for MCL and MMSE estimates of AR coefficients for 2nd order AR process with $\rho = .8$ and $\theta = 0$ radians.

As can be seen from the tables, the proposed approach yields superior estimates and outperforms the MCL method for these pole positions.

To visualize performance, scatter plots for the MCL and MMSE estimates are shown in Figures 1-3. The values of ρ and θ are the same as in Tables 1-3, respectively. The scatter plots show that the Bayesian approach clusters the estimates closer to the true value of the AR coefficients which are located at the crosshairs. Figure 1 indicates a bias in the Bayesian method, which is reasonable since the coefficients lie on the boundary of Ω and no estimates can occur outside this region. In Figures 2 and 3, the true values of the coefficients are well within Ω and the plots clearly indicate the benefit of computing MMSE estimates regardless of the stationary constraint.

	MCL	MMSE
ϵ_{a_1}	.347	.281
ϵ_{a_2}	.353	.113

Table 2: RMS errors for MCL and MMSE estimates of AR coefficients for 2nd order AR process with $\rho = .8$ and $\theta = \pi/8$ radians.

	MCL	MMSE
ϵ_{a_1}	.299	.273
ϵ_{a_2}	.306	.167

Table 3: RMS errors for MCL and MMSE estimates of AR coefficients for 2nd order AR process with $\rho = .8$ and $\theta = \pi/2$ radians.

The simulations were run for many additional pole positions with the results shown in Figures 4-7. In these plots, the combined RMS error of the AR coefficient estimates are shown as a function of θ for fixed ρ. Figure 4 is the performance with $\rho = .2$. Since the poles are close to zero, in the complex plane, the process looks quite like the driving noise sequence which helps to explain why the performance of the MCL and MMSE are nearly equivalent. As ρ approaches 1, which implies the poles approach the unit circle, the improvement in performance that is obtained by using the Bayesian approach becomes more apparent. For $\rho = .8$, in Figure 7, the MMSE estimates are clearly superior with at least a 50% improvement in RMS error. As θ approaches zero, and the poles become close together, the MCL performance degrades substantially while the MMSE actually improves slightly. This is particularly interesting because in many applications the poles of the AR process are located near each other and these results indicate that the Bayesian approach will enjoy a certain degree of immunity.

To examine the performance as the sample size increases, an experiment was performed with the number of samples, N, equal to 25. The RMS curves are shown in Figure 8 for the case of $\rho = .8$. Clearly, the MMSE estimates are still better than the MCL though the margin of improvement is less than in the case of $N = 10$. This is expected since as the number of samples grows, the MCL becomes a better approximation to the actual MMSE estimate. These simulation results clearly illustrate the benefit of using a Bayesian methodology as well as provide evidence of the efficacy of the importance sampling technique.

6 Conclusions

In this paper, we have introduced a Bayesian solution to the problem of AR signal modelling. By utilizing the exact likelihood expression and formulating the theoretical covariances in terms of the AR coefficients, the posterior density was constructed using noninformative priors for both the AR coefficients and the signal variance, which was marginalized analytically. The resulting posterior density can then be used to examine many signal descriptors or characteristics.

To perform the required integrations and generate MMSE estimates, a highly efficient numerical technique, importance sampling, was introduced and how it can be applied to the AR parameter estimation problem was discussed. The method guarantees consistent estimates of the necessary integrals and, further, avoids the pitfalls associated with multidimensional optimization.

In addition, simulation experiments were performed where the MMSE and MCL estimates of the AR coefficients were compared. The results indicate the improvement that can be attained by using a Bayesian approach and in all cases examined the MMSE estimates were superior. In the future, it will be interesting to examine the problems of prediction and spectral estimation with the AR model using the proposed approach and to compare performance with classical methods. Also, since the importance sampling procedure has clear benefits as compared to optimization methods, it will be worthwhile to apply this technique, with the Bayesian philosophy, to other signal processing problems, such as detection and estimation of sinusoidal signals.

References

[1] G. Bodenstein and H. M. Praetorius, "Feature extraction from the electroencepalogram by adaptive segmentation," *Proc. of the IEEE*, Vol 65, pp 642-652, May, 1977.

[2] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control.* San Francisco, Ca.: Holden-Day, 1970.

[3] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley, 1973.

[4] P. M. T. Broersen, "The prediction error of autoregressive small sample models," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 38, pp 858-860, May, 1990.

[5] P. M. Djurić and S. M. Kay, "Order selection of autoregressive models," *IEEE Trans. Signal Processing*, Vol. 40, pp 2829-2833, Nov. 1992.

[6] F. James, "Monte Carlo theory and practice," *Rep. Prog. Phys.*, Vol. 43, pp 1145-1175, 1980.

[7] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, New York, NY: John Wiley & Sons, 1986.

[8] S. K. Katsikas, S. D. Likothanassis and D. G. Lainiotis, "AR model identification with unknown process order," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 38, pp 872-876, May, 1990.

[9] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Englewood Cliffs, NJ: Prentice Hall, 1988.

[10] S. M. Kay and V. Nagesha, "Extraction of periodic signals in colored noise," *Proc. ICASSP*, Vol. 5, pp 281-284, 1993.

[11] J. Makhoul, "Linear prediction: A tutorial review," *Proc. of the IEEE*, Vol. 63, pp. 561-580, April, 1975.

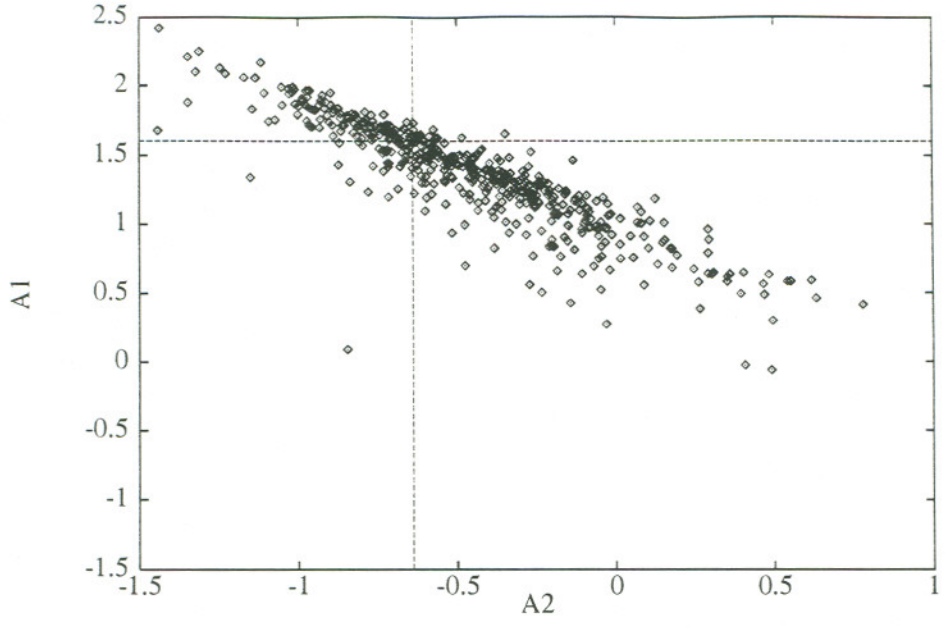
[12] M. S. Oh and J. O. Berger, "Integration of multimodal functions by Monte Carlo importance sampling," *Journal of American Statistical Association*, Vol. 88, pp. 450-456, June 1993.

[13] M. B. Priestley, *Spectral Analysis and Time Series, Volume 1*, New York, NY: Academic Press, 1981.

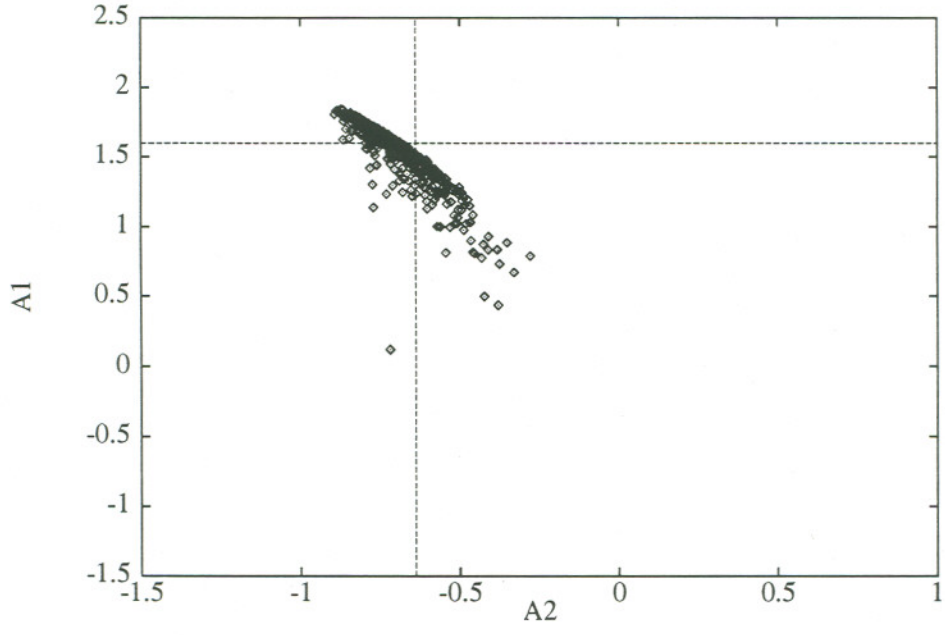
[14] A. F. M. Smith, "Bayesian computational methods," *Phil. Trans. R. Soc. Lond. A*, Vol. 337, pp. 369-386, 1991.

[15] A. H. Wang and R. L. Klein, "Optimal quadrature formula nonlinear estimators," *Information Sciences*, Vol. 16, pp. 169-184, 1978.

[16] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, Malabar, Fla.: Krieger, 1987.

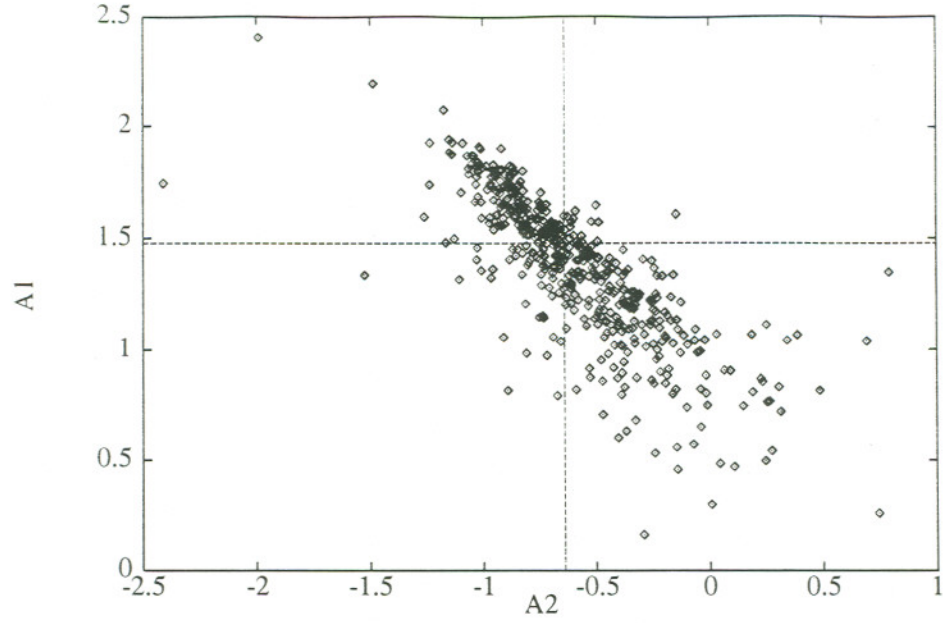


(a)

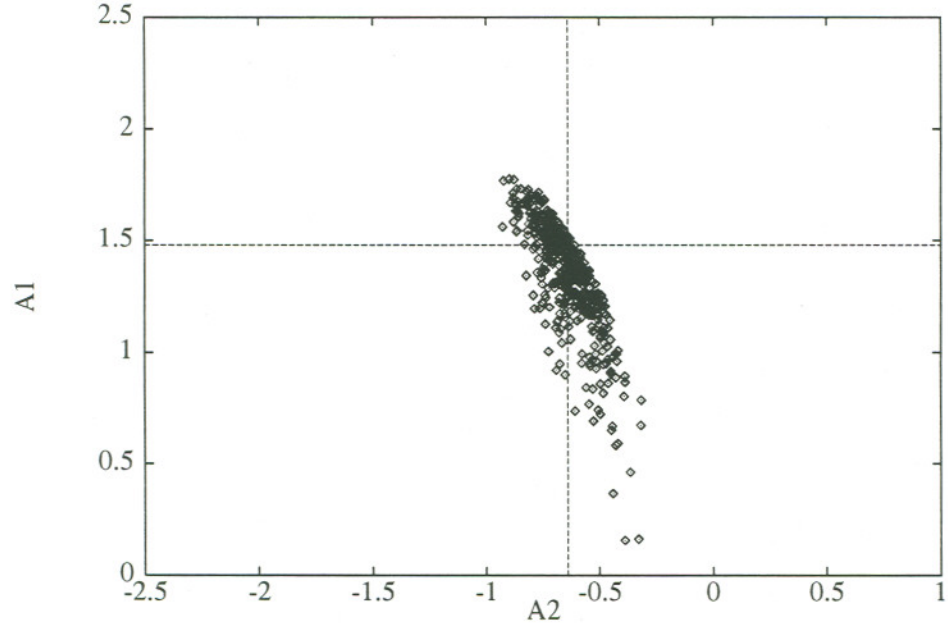


(b)

Figure 1: Scatter plots of (a) MCL coefficient estimates and (b) MMSE estimates of AR coefficients for 2nd order AR process with $\rho = .8$ and $\theta = 0$ radians. Crosshairs indicate true values of a_1 and a_2.

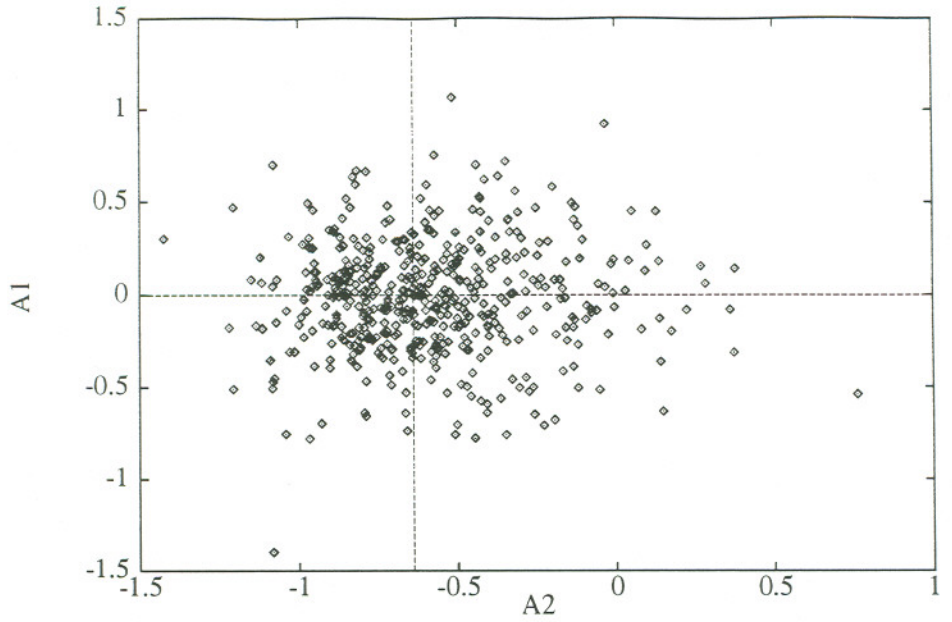


(a)

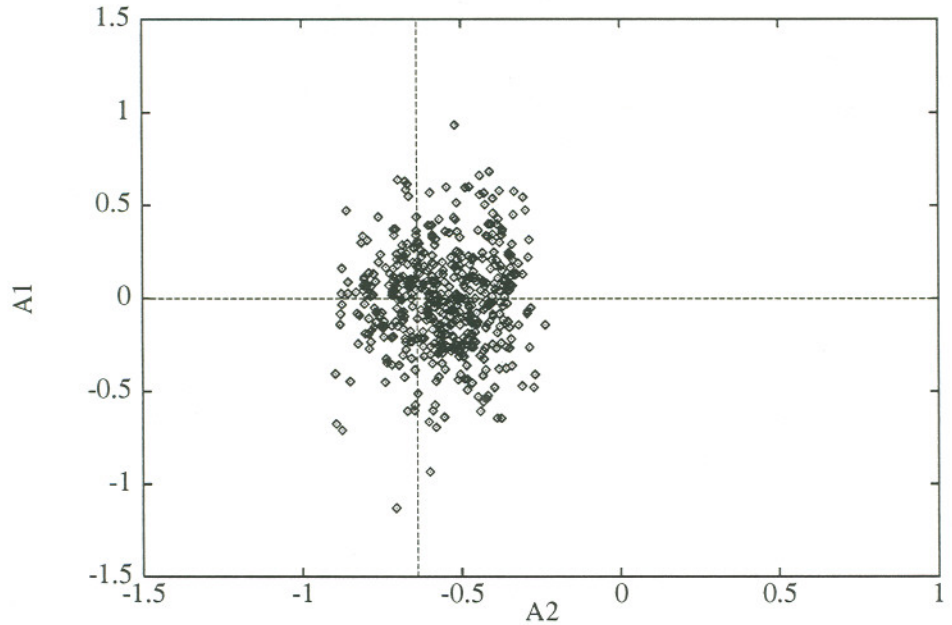


(b)

Figure 2: Scatter plots of (a) MCL coefficient estimates and (b) MMSE estimates of AR coefficients for 2nd order AR process with $\rho = .8$ and $\theta = \pi/8$ radians. Crosshairs indicate true values of a_1 and a_2.



(a)



(b)

Figure 3: Scatter plots of (a) MCL coefficient estimates and (b) MMSE estimates of AR coefficients for 2nd order AR process with $\rho = .8$ and $\theta = \pi/2$ radians. Crosshairs indicate true values of a_1 and a_2.

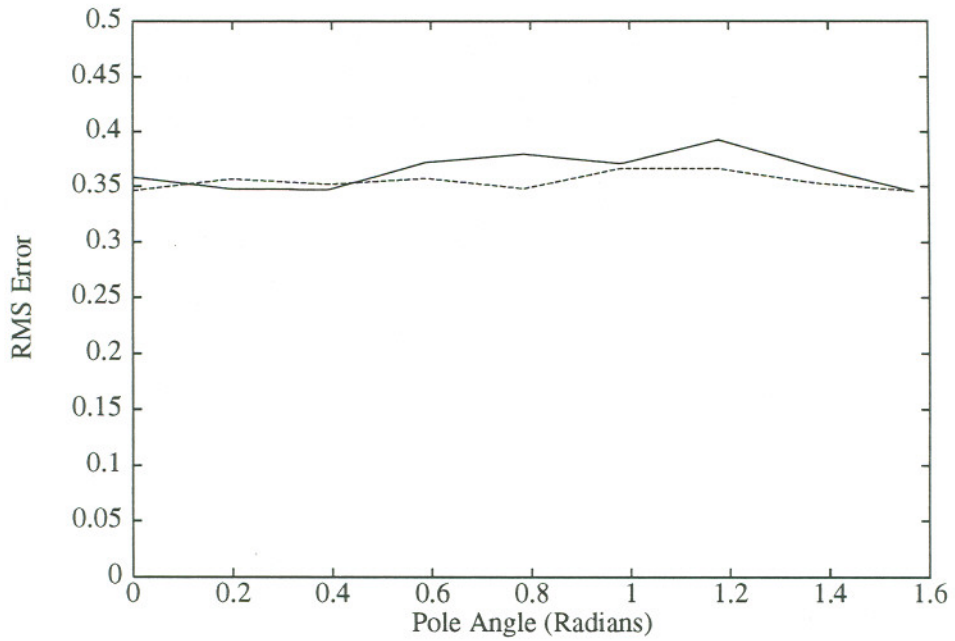


Figure 4: RMS error of MCL (solid line) and MMSE (dashed line) estimates of AR coefficients as a function of pole angle, θ, for 2nd order AR process with $\rho = .2$.

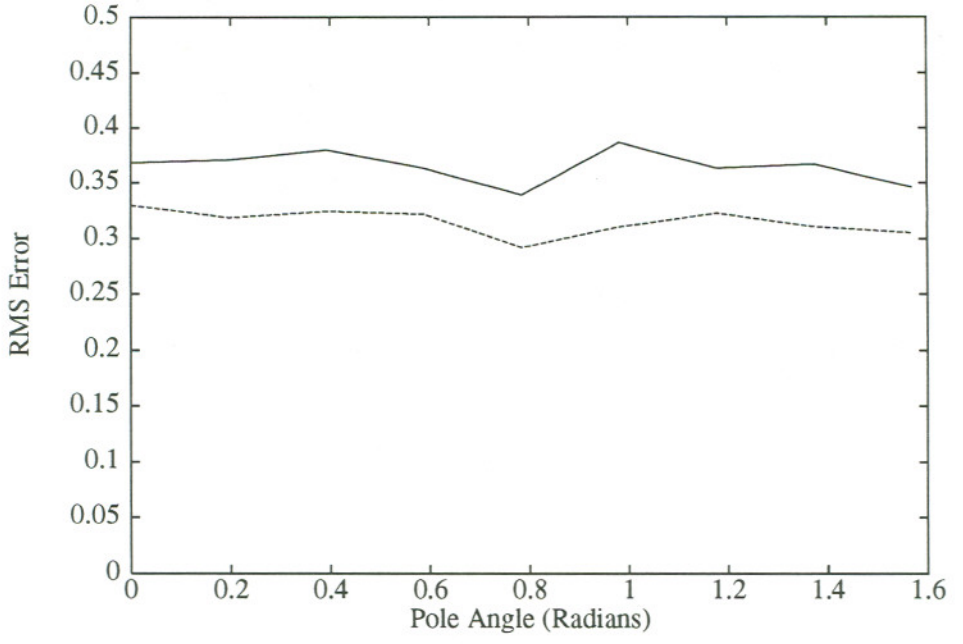


Figure 5: RMS error of MCL (solid line) and MMSE (dashed line) estimates of AR coefficients as a function of pole angle, θ, for 2nd order AR process with $\rho = .4$.

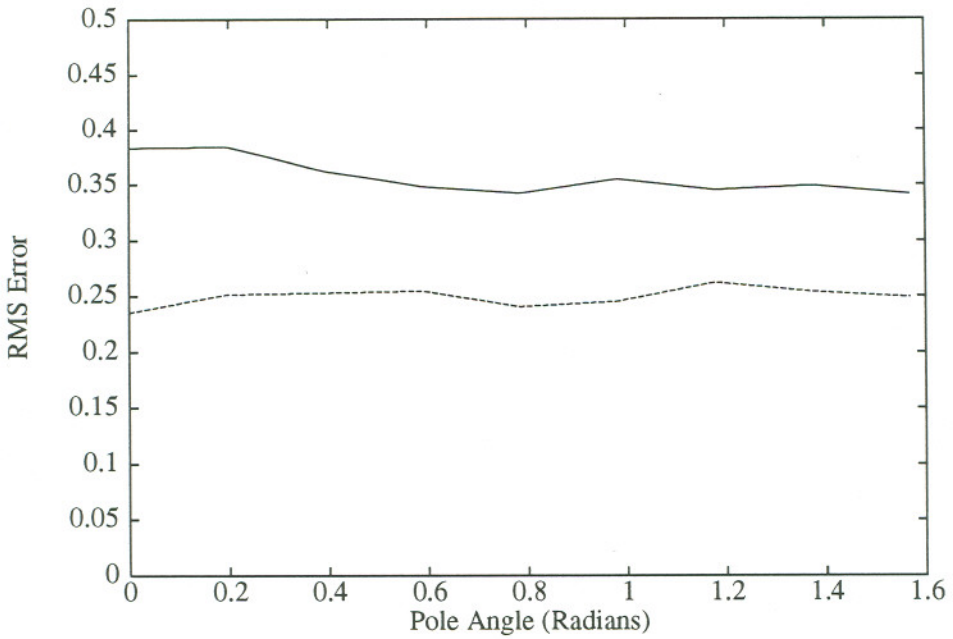


Figure 6: RMS error of MCL (solid line) and MMSE (dashed line) estimates of AR coefficients as a function of pole angle, θ, for 2nd order AR process with $\rho = .6$.

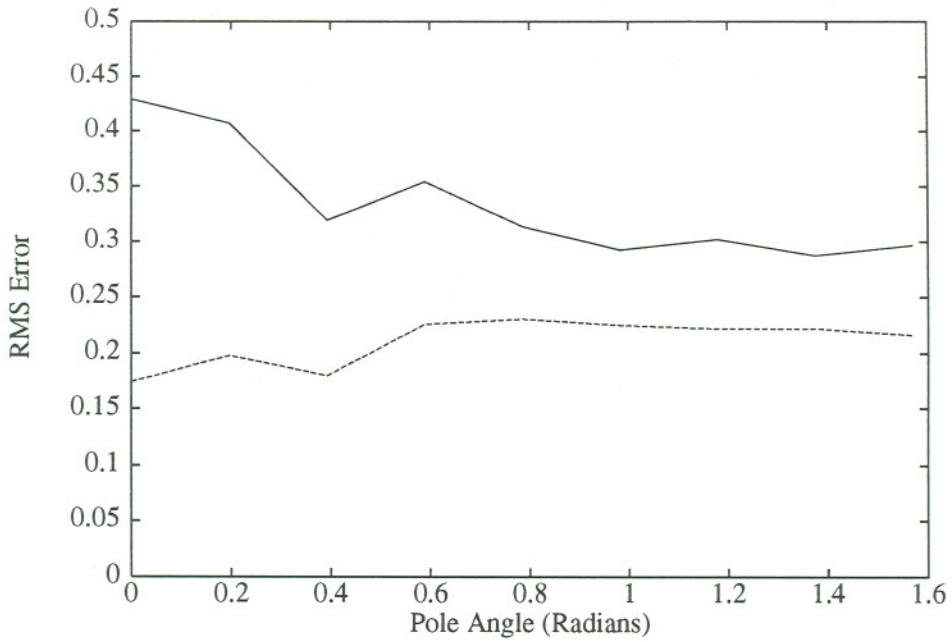


Figure 7: RMS error of MCL (solid line) and MMSE (dashed line) estimates of AR coefficients as a function of pole angle, θ, for 2nd order AR process with $\rho = .8$.

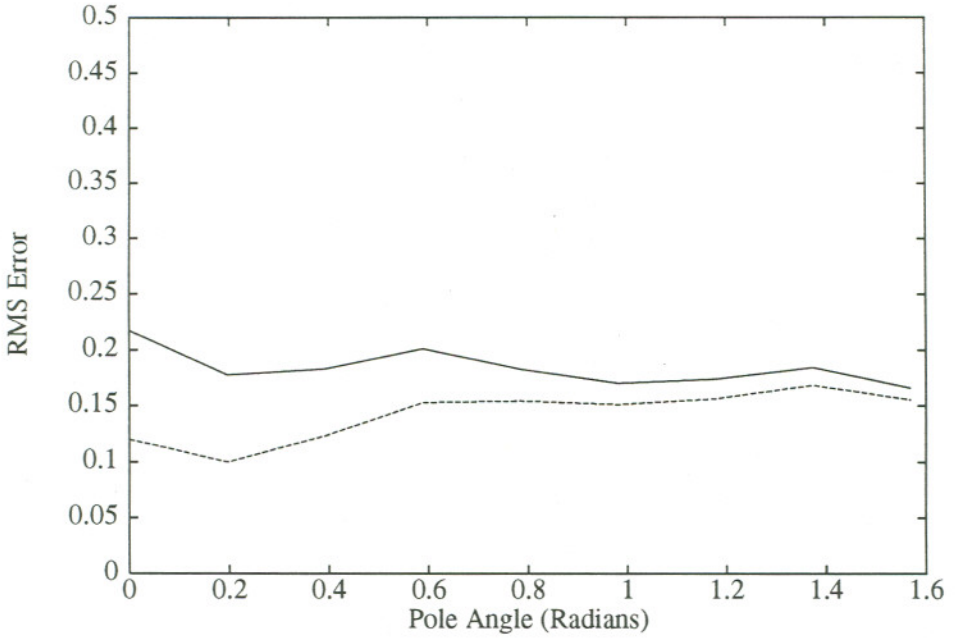


Figure 8: RMS error of MCL (solid line) and MMSE (dashed line) estimates of AR coefficients as a function of pole angle, θ, for 2nd order AR process with $\rho = .8$ and $N = 25$.