International Scholarly Research Network
ISRN Signal Processing
Volume 2011, Article ID 138683, 10 pages
doi:10.5402/2011/138683

Research Article

Linear Estimation of Stationary Autoregressive Processes

Reza Dianat[1] and Farokh Marvasti[2]

[1] *Engineering Department, Persian Gulf University, Davvas, 75169-13798 Bushehr Port, Iran*
[2] *Advanced Communication Research Center, Sharif University of Technology, P.O. Box 11356-11155, Tehran, Iran*

Correspondence should be addressed to Reza Dianat, dianat@pgu.ac.ir

Received 1 December 2010; Accepted 12 January 2011

Academic Editors: K. M. Prabhu and A. Tefas

Consider a sequence of an mth-order Autoregressive (AR) stationary discrete-time process and assume that at least $m-1$ consecutive neighboring samples of an unknown sample are available. It is not important that the neighbors are from one side or are from both the left and right sides. In this paper, we find explicit solutions for the optimal linear estimation of the unknown sample in terms of the neighbors. We write the estimation errors as the linear combination of innovation noises. We also calculate the corresponding mean square errors (MSE). To the best of our knowledge, there is no explicit solution for this problem. The known solutions are the implicit ones through orthogonality equations. Also, there are no explicit solutions when fewer than $m-1$ samples are available. The order of the process (m) and the feedback coefficients are assumed to be known.

1. Introduction

Estimation has many applications in different areas including compression and equalization [1, 2]. The linear estimation is more common due to its mathematical simplicity. The optimal linear estimation of a random variable x in terms of y_1, y_2, and y_n is the following linear combination

$$\hat{x} \triangleq \hat{E}\{x \mid y_1, y_2, \ldots, y_n\} = \sum_{i=1}^{n} A_i y_i, \tag{1}$$

where the coefficients A_i must be chosen to minimize the MSE $E\{(x-\hat{x})^2\}$ and $E\{\cdot\}$ stands for the expected value. To minimize the MSE, we must choose A_i's to satisfy the orthogonality principle as follows:

$$E\{(x-\hat{x})y_i\} = 0, \quad i = 1, 2, \ldots, n. \tag{2}$$

We also write the above condition as

$$x - \hat{x} \perp y_i, \quad i = 1, 2, \ldots, n. \tag{3}$$

Therefore, in the optimal linear estimation, we search for the coefficients such that the error is orthogonal to the data.

A common model for many signals including image, speech, and biological signals is the AR model [1, 3–5]. This model has applications in different areas including detection [6, 7], traffic modeling [8], channel modeling [9], and forecasting [10]. An AR process is the output of an all-pole causal filter whose input is a white sequence called innovation noise [11]. We introduce another model for the process using an all-pole anticausal filter as well. The optimal linear estimation of an AR process is accomplished through the recursive solution of Yule-Walker (YW) equations using Levinson-Durbin algorithm [12]. This solution is recursive and implicit. As we will see in some cases the equation coefficients do not form a Toeplitz matrix and we cannot enjoy the complexity reduction advantage of Levinson algorithm.

To the best of our knowledge, there is no explicit solution for YW equations. Most of the focus of researchers is on model parameters estimation from observations. When researchers arrive at YW equations, they stop, since they consider the solution as known through Levinson recursion. Broersen in his method for autocorrelation function estimation form observations points to YW equations and mainly concentrates on bias reduction in estimation using finite set of observations [13, 14]. He does not attempt to find the solution for YW equations. Fattah et al. try to estimate the autocorrelation function of an ARMA model from noisy data; they again refer to YW equation set and its solution using matrix inversion and no explicit solution is proposed [15]. Xia and Kamel propose an optimization method to estimate AR model parameters from noisy data [16]. Noise is not necessarily Gaussian. The method finds a minimum for a cost function and exploits a neural-network algorithm. Again, the explicit solution of the orthogonality equations is not the goal of the paper. Hsiao proposes an algorithm to estimate the parameters of a time-varying AR system [17]. He considers the feedback coefficients of a time-varying AR process as random variables. The proposed algorithm maximizes a posteriori probabilities conditioned on the data. The recursive algorithm is compared to Monte Carlo simulation in terms of accuracy and complexity. In this paper, the aim is parameter estimation from data and not the analytic solution of orthogonality equations. In [18], a sequence of Gaussian AR vector is considered. As the sequence elements are vectors rather than scalars, the AR model is defined by matrix feedback coefficients rather than scalar feedback coefficients. The estimation here is more complex, and some independence conditions are assumed. The method is based on convex optimization, and no exact answer can be provided. Mahmoudi and Karimi propose an LS-based method to estimate AR parameters from noisy data [19]. The method exploites YW equations, but this method also does not provide the explicit solution to the equations. Another LS-based estimation method can be seen in [20].

As mentioned above, we could not see the final solution to YW orthogonality equations in the literature. In this work we have derived explicit solutions for orthogonality equations for different cases. Consider a stationary mth order AR process. The order and the feedback coefficients of the process are assumed to be known, and the model parameter estimation is out of the scope of this paper. The main goal of this paper is finding the solution for the orthogonality equations. We will find the the optimal linear estimation of a sample in terms of the neighbors where at least $m-1$ consecutive neighbors are available. The consecutive neighbors include the situations where all the $m-1$ or more neighbors are in one side, or some of them are left neighbors and the others are right neighbors. We will show that no more that m consecutive neighbors in each side are needed. Our approach is to find orthogonal estimation errors that are linear combinations of data. We use the well-known causal LTI AR model as well as our anticausal model to form orthogonal errors. The errors are formed as a linear combination of causal and anticausal innovation (process) noises. Beginning from suitable errors that are both orthogonal to the data and are linear combination of data, we arrive at linear estimations. We seek LTI system approach rather than trying to directly solve the orthogonality equations. The results of this paper for different cases can be important in situations where the equation matrices are ill-posed and the matrix inversion and other recursive algorithms become unstable.

This paper is organized as follows. In Section 2, the causal model is reviewed and the anticausal model is introduced. In Section 3, we review the forward prediction problem. We state the problem symmetries in Section 4. We see how we can use the similarities between two problems to exploit the solution of one problem to find the solution of the other problem. In Section 5, we extract a number of relations for cross-correlation functions that will be used later. We find the interpolation formulae when infinite data are available in Section 6. We find the prediction and interpolation with finite data in Sections 7 and 8, respectively. In Section 9, we present a detailed example to show that our relations and the matrix solution of the orthogonality principle result in the same coefficients. Finally, we conclude the work in Section 10.

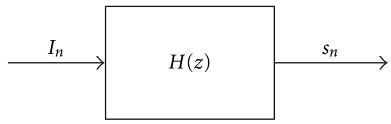

Figure 1: The causal model.

2. Causal and Anticausal Models

A discrete-time stationary AR process s_n of order m is modeled as follows.

$$s_n + a_1 s_{n-1} + a_2 s_{n-2} + \cdots + a_m s_{n-m} = I_n, \quad n \in \mathbb{Z}. \tag{4}$$

The above equation is meant for a causal LTI system. I_n, the input of the system, is called the innovation noise and is a stationary white sequence with the zero expected value, that is, $E\{I_n I_k\} = \sigma^2 \delta[n-k]$ and $E\{I_n\} = 0$, where σ is a positive constant. $\delta[0] = 1$ and $\delta[i] = 0$ elsewhere. The system is causal. Therefore s_n, the output of the system in the time index n, is a linear combination of the inputs in the time index n and before. So, we can write

$$s_n = h_0 I_n + h_1 I_{n-1} + h_2 I_{n-2} + \cdots = \sum_{i=0}^{\infty} h_i I_{n-i}. \tag{5}$$

In the above equation, h_n is the impulse response of the system. Assuming the causal system model, we have $h_n = 0$ for $n < 0$. Paying attention to the whiteness of the sequence $\{I_n\}$ and from (5) we get the following result.

$$I_{n+k} \perp s_n, \quad k > 0, \; n \in \mathbb{Z}. \tag{6}$$

Figure 1 is the causal model of the AR process. $H(z)$ is the Z-transform of h_n, which is defined as

$$H(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k}. \tag{7}$$

For the system defined by (4), we have

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + \cdots + a_m z^{-m}}. \tag{8}$$

Assuming a stable causal system, we conclude that the roots of $A(z) = 0$ must be inside the unit circle $|z| = 1$. The power spectral density function (PSDF) of a process is the Z-transform of its autocorrelation function. The PSDF of s_n is [11]

$$S_s(z) = S_I(z) H(z) H(z^{-1}) = \sigma^2 H(z) H(z^{-1}). \tag{9}$$

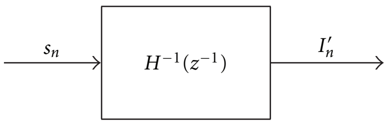

Figure 2: Generation of another innovation noise.

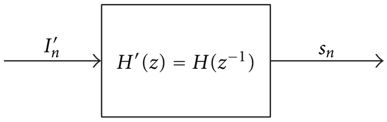

Figure 3: Anticausal model.

In the above equation $S_s(z)$ is the PSDF of s_n and $S_I(z)$ is the PSDF of I_n.

We now present the anticausal model. If we apply the sequence s_n to an LTI system with the transfer function $H^{-1}(z^{-1})$, we get another innovation noise called I'_n. Figure 2 demonstrates the generation of the new innovation noise. To see the whiteness of the sequence I'_n, note that the PSDF of I'_n by using Figure 2 and (9) is as follows.

$$\begin{aligned} S_{I'}(z) &= S_s(z)H^{-1}(z^{-1})H^{-1}(z) \\ &= \sigma^2 H(z)H(z^{-1})H^{-1}(z^{-1})H^{-1}(z) = \sigma^2. \end{aligned} \tag{10}$$

Equivalently we can apply I'_n to the inverse system with the transfer function $H'(z) = H(z^{-1})$ to get s_n. The generation of s_n from I'_n is depicted in Figure 3.

We have

$$H'(z) = H(z^{-1}) = \frac{1}{A(z^{-1})}. \tag{11}$$

Therefore $h'_n = h_{-n}$. Noting that $h_n = 0$ for $n < 0$, we see that $h'_n = 0$ for $n > 0$. Also note that the roots of $A(z^{-1}) = 0$ are outside the unit circle, as we had the roots of $A(z) = 0$ inside the unit circle. Regarding these points, we know that the system with the transfer function $H'(z)$ is stable and anticausal. We have

$$H'(z) = \frac{1}{1 + a_1 z + a_2 z^2 + \cdots + a_m z^m}. \tag{12}$$

Using the above equation and Figure 3, we get

$$s_n + a_1 s_{n+1} + a_2 s_{n+2} + \cdots + a_m s_{n+m} = I'_n, \quad n \in \mathbb{Z}. \tag{13}$$

Also, note that

$$\begin{aligned} s_n &= \sum_{i=-\infty}^{\infty} h'_i I'_{n-i} = h'_0 I'_n + h'_{-1} I'_{n+1} + h'_{-2} I'_{n+2} + \cdots \\ &= h_0 I'_n + h_1 I'_{n+1} + \cdots = \sum_{i=0}^{\infty} h_i I'_{n+i}. \end{aligned} \tag{14}$$

From (14) and Figure 3, we see that s_n is a linear combination of I'_n and the inputs after that. The whiteness of the sequence $\{I'_n\}$ gives then

$$I'_{n-k} \perp s_n, \quad n \in \mathbb{Z},\ k > 0. \tag{15}$$

3. Forward Prediction

Forward prediction can be accomplished by using the whitening filter [11]. The data are whitened, and we use the equivalent white data to achieve the prediction. As an example, consider the 1-step forward prediction of s_n. It is seen that s_n is estimated as

$$\begin{aligned} \hat{s}_n &= \hat{E}\{s_n \mid s_{n-k}, k > 0\} = -\sum_{k=1}^{m} a_k s_{n-k} \\ &= -a_1 s_{n-1} - a_2 s_{n-2} - \cdots - a_m s_{n-m}. \end{aligned} \tag{16}$$

It can bee seen from (4) that the error $s_n - \hat{s}_n$ is equal to I_n and therefore, from (6), it is orthogonal to s_{n-k} for $k > 0$. It proves the optimality of (16).

The 2-step prediction can be done as [11]

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n-k}, k \ge 2\} = -a_1 \hat{s}_{n-1} - \sum_{i=2}^{m} a_i s_{n-i}. \tag{17}$$

In the above equation, $\hat{s}_{n-1}$ is the prediction of s_{n-1} from its previous data (1-step prediction) and is obtained by replacing n by $n-1$ in (16).

$$\begin{aligned} \hat{s}_{n-1} &= \hat{E}\{s_{n-1} \mid s_{n-k},\ k \ge 2\} = -\sum_{k=1}^{m} a_k s_{n-k-1} \\ &= -a_1 s_{n-2} - a_2 s_{n-3} - \cdots - a_m s_{n-m-1}. \end{aligned} \tag{18}$$

From (17), (18), and (4), the estimation error is

$$\begin{aligned} e_n &= s_n + \sum_{i=2}^{m} a_i s_{n-i} - a_1 \sum_{k=1}^{m} a_k s_{n-k-1} \\ &= I_n - a_1 s_{n-1} - a_1 \sum_{k=1}^{m} a_k s_{n-k-1} = I_n - a_1 I_{n-1}. \end{aligned} \tag{19}$$

From (6), it is clear that I_n and I_{n-1} are orthogonal to s_{n-k} for $k \ge 2$. It proves the optimality of (17).

The higher-order predictions can be obtained in the same manner. As the final example of this section, consider the 3-step forward prediction that is accomplished as follows.

$$\begin{aligned} \hat{s}_n &= \hat{E}\{s_n \mid s_{n-k}, k \ge 3\} \\ &= -a_1 \hat{s}_{n-1} - a_2 \hat{s}_{n-2} - \sum_{k=3}^{m} a_k s_{n-k}. \end{aligned} \tag{20}$$

In the above equation, $\hat{s}_{n-1}$ and $\hat{s}_{n-2}$ are the 2-step and 1-step predictions of s_{n-1} and s_{n-2}, respectively, and are obtained from (17) and (16). The error is

$$\begin{aligned} e_n &= s_n + a_1 \hat{s}_{n-1} + a_2 \hat{s}_{n-2} + \sum_{i=3}^{m} a_i s_{n-i} \\ &= I_n - a_1 s_{n-1} - a_2 s_{n-2} + a_1 \hat{s}_{n-1} + a_2 \hat{s}_{n-2} \\ &= I_n - a_1 (s_{n-1} - \hat{s}_{n-1}) - a_2 (s_{n-2} - \hat{s}_{n-2}) \\ &= I_n - a_1 (I_{n-1} - a_1 I_{n-2}) - a_2 I_{n-2}. \end{aligned} \tag{21}$$

4. The Problem Symmetries

Consider the following linear interpolation of s_n from the data around it:

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n-k_1}, s_{n-k_1+1}, \dots, s_{n-1}, s_{n+1}, \dots, s_{n+k_2}\} \\ &= a'_{-k_1} s_{n-k_1} + a'_{-k_1+1} s_{n-k_1+1} + \cdots + a'_{k_2} s_{n+k_2}.\end{aligned} \tag{22}$$

The orthogonality principle gives

$$\begin{aligned}&E\Big\{\Big(s_n - a'_{-k_1} s_{n-k_1} - a'_{-k_1+1} s_{n-k_1+1} - \cdots - a'_{k_2} s_{n+k_2}\Big) s_{n+i}\Big\} \\ &\quad = 0, \quad i = -k_1, -k_1+1, \dots, k_2, \qquad i \neq 0.\end{aligned} \tag{23}$$

The above equations become

$$\begin{aligned}&R_s[i+k_1]a'_{-k_1} + R_s[i+k_1-1]a'_{-k_1+1} + \cdots + R_s[i-k_2]a'_{k_2} \\ &\quad = R_s[i], \quad i = -k_1, -k_1+1, \dots, k_2, \qquad i \neq 0.\end{aligned} \tag{24}$$

In the above equations, $R_s[i] = E\{s_n s_{n-i}\}$.

Now, consider the following estimation.

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n-k_2}, s_{n-k_2+1}, \dots, s_{n-1}, s_{n+1}, \dots, s_{n+k_1}\} \\ &= a''_{k_1} s_{n+k_1} + a''_{k_1-1} s_{n+k_1-1} + \cdots + a''_{-k_2} s_{n-k_2}.\end{aligned} \tag{25}$$

The orthogonality of error to the data gives

$$\begin{aligned}&E\Big\{\Big(s_n - a''_{k_1} s_{n+k_1} - a''_{k_1-1} s_{n+k_1-1} - \cdots - a''_{-k_2} s_{n-k_2}\Big) s_{n+i}\Big\} \\ &\quad = 0, \quad i = k_1, k_1-1, \dots, -k_2, \qquad i \neq 0.\end{aligned} \tag{26}$$

They become

$$\begin{aligned}&R_s[i-k_1]a''_{k_1} + R_s[i-k_1+1]a''_{k_1-1} + \cdots + R_s[i+k_2]a''_{k_2} \\ &\quad = R_s[i], \quad i = k_1, k_1-1, \dots, -k_2, \qquad i \neq 0.\end{aligned} \tag{27}$$

Regarding that the $R_s[\cdot]$ is an even function, we notice that the set of equations (24) and the set of equations (27) are exactly the same. Therefore,

$$a'_{-k_1} = a''_{k_1}, a'_{-k_1+1} = a''_{k_1-1}, \dots, a'_{k_2} = a''_{-k_2}. \tag{28}$$

As an example, consider the following backward prediction.

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n+k}, k > 0\}. \tag{29}$$

Using (16) and the symmetry, we get

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n+k},\ k > 0\} = -\sum_{k=1}^{m} a_k s_{n+k} \\ &= -a_1 s_{n+1} - a_2 s_{n+2} - \cdots - a_m s_{n+m}.\end{aligned} \tag{30}$$

The validity of the solution can also be confirmed as from (13), it is seen that the estimation error is

$$s_n + a_1 s_{n+1} + a_2 s_{n+2} + \cdots + a_m s_{n+m} = I'_n. \tag{31}$$

Using (15), it is clear that the error is orthogonal to the data. It proves the optimality of (30).

5. Cross-Correlation Functions

In this section, we derive a number of properties for the cross-correlations between innovation noises and the AR process. We will exploit these properties to prove our solutions.

We define $R_{sI}[k] = E\{s_n I_{n-k}\}$ and $R_{I's}[k] = E\{I'_n s_{n-k}\}$. The first simple property follows from (6) and (15) as follows.

$$R_{sI}[k] = R_{I's}[k] = 0, \quad k < 0. \tag{32}$$

Now, consider Figure 1. In this figure I_n is the input and s_n is the output. The impulse response of system is $h_n \triangleq h[n]$. Therefore, we have [11]

$$R_{sI}[k] = R_I[k] * h[k] = \sigma^2 \delta[k] * h[k] = \sigma^2 h[k]. \tag{33}$$

In this equation, $R_I[k] = E\{I_n I_{n-k}\}$ and the "$*$" operator is the discrete convolution. Taking the Z-transform from both sides of (33) and using (8), we get

$$S_{sI}(z) = \sigma^2 H(z) = \frac{\sigma^2}{1 + a_1 z^{-1} + \cdots + a_m z^{-m}}. \tag{34}$$

Or equivalently

$$S_{sI}(z)\left(1 + a_1 z^{-1} + \cdots + a_m z^{-m}\right) = \sigma^2. \tag{35}$$

Taking inverse Z-transform from this equation, we have

$$R_{sI}[k] + a_1 R_{sI}[k-1] + \cdots + a_m R_{sI}[k-m] = \sigma^2 \delta[k]. \tag{36}$$

The right side of (36) is zero for $k \neq 0$.

Referring to Figure 3, we have [11]

$$\begin{aligned}R_{I's}[k] &= R_{I'}[k] * h'[-k] = \sigma^2 \delta[k] * h'[-k] \\ &= \sigma^2 h'[-k] = \sigma^2 h[k].\end{aligned} \tag{37}$$

Again, we conclude that

$$R_{I's}[k] + a_1 R_{I's}[k-1] + \cdots + a_m R_{I's}[k-m] = \sigma^2 \delta[k]. \tag{38}$$

6. Interpolation Using an Infinite Set of Data

In this section, we assume that infinite number of data are available. However, we will see that only a finite number of data are sufficient.

6.1. Infinite Data on the Left Side. We want to obtain the following estimation.

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n+i},\ i \leq k_1,\ i \neq 0\}. \tag{39}$$

k_1 is a positive integer constant not greater than m. There are k_1 data available on the right side of s_n and infinite data on the left side. Define $a_0 \triangleq 1$. We are going to prove the following:

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n+i},\ i \le k_1, i \neq 0\}$$
$$= -\frac{1}{\sum_{k=0}^{k_1} a_k^2}\left(\sum_{k=1}^{k_1}\left(\sum_{p=0}^{k_1-k} a_p a_{p+k}\right)s_{n+k} + \sum_{k=1}^{m-k_1}\left(\sum_{p=0}^{k_1} a_p a_{p+k}\right)s_{n-k} + \sum_{k=m-k_1+1}^{m}\left(\sum_{p=0}^{m-k} a_p a_{p+k}\right)s_{n-k}\right). \tag{40}$$

Observe from (40) that although there are infinite data on the left side of s_n, only m data s_{n-1} to s_{n-m} participate in the estimation. Indeed, (40) is the optimal linear estimation solution for $\hat{s}_n = \hat{E}\{s_n \mid s_{n+i},\ -k_2 \le i \le k_1,\ i \neq 0\}$, where k_2 can be any integer greater than or equal to m.

To prove the optimality of (40), we must show that the estimation error is orthogonal to the data. Firstly the estimation error can be calculated by inserting $\hat{s}_n$ from (40) in $e_n = s_n - \hat{s}_n$. Secondly, by extending the innovation noises using (4) we can confirm that

$$e_n = s_n - \hat{s}_n = \frac{1}{\sum_{k=0}^{k_1} a_k^2}\left(I_n + a_1 I_{n+1} + \cdots + a_{k_1} I_{n+k_1}\right). \tag{41}$$

Indeed, we have obtained (40) from (41). The motivation is that the estimation error has to possess two essential conditions: (1) it must be orthogonal to the data and (2) it must be only a linear combination of the data and the variable to be estimated. It remains to prove that the right side of (41) is orthogonal to the data.

Using (6), it is quite clear that I_n to I_{n+k_1} are orthogonal to s_{n-k} for $k > 0$, and so is e_n in (41). Further, we have

$$\begin{aligned} &E\{s_{n+i}(I_n + a_1 I_{n+1} + \cdots + a_{k_1} I_{n+k_1})\} \\ &\quad = R_{sI}[i] + a_1 R_{sI}[i-1] + \cdots + a_{k_1} R_{sI}[i-k_1] \\ &\quad = R_{sI}[i] + a_1 R_{sI}[i-1] + \cdots + a_i R_{sI}[0], \quad 1 \le i \le k_1. \end{aligned} \tag{42}$$

The last equation of (42) is justified as we have $R_{sI}[k] = 0$ for $k < 0$ from (32). Using (32), (36), (41), and (42) it is seen that

$$E\{(I_n + a_1 I_{n+1} + \cdots + a_{k_1} I_{n+k_1})s_{n+i}\} = 0, \quad 1 \le i \le k_1. \tag{43}$$

This completes the proof.

The MSE is

$$\begin{aligned} E\{e_n^2\} &= \frac{1}{\left(\sum_{k=0}^{k_1} a_k^2\right)^2} \cdot E\left\{(I_n + a_1 I_{n+1} + \cdots + a_{k_1} I_{n+k_1})^2\right\} \\ &= \frac{1}{\left(\sum_{k=0}^{k_1} a_k^2\right)^2} \cdot \left(E\{I_n^2\} + a_1^2 E\{I_{n+1}^2\} + \cdots + a_{k_1}^2 E\left\{I_{n+k_1}^2\right\}\right) \\ &= \frac{1}{\left(\sum_{k=0}^{k_1} a_k^2\right)^2} \cdot \left(\sigma^2 + a_1^2\sigma^2 + \cdots + a_{k_1}^2\sigma^2\right). \end{aligned} \tag{44}$$

Therefore,

$$E\{e_n^2\} = \frac{\sigma^2}{\sum_{k=0}^{k_1} a_k^2}. \tag{45}$$

6.2. Infinite Data on the Right Side. By symmetry, and replacing s_{n-k} by s_{n+k} in (40), the following estimation is derived.

$$\begin{aligned} \hat{s}_n &= \hat{E}\{s_n \mid s_{n-i},\ i \le k_1,\ i \neq 0\} \\ &= -\frac{1}{\sum_{k=0}^{k_1} a_k^2}\left(\sum_{k=1}^{k_1}\left(\sum_{p=0}^{k_1-k} a_p a_{p+k}\right)s_{n-k} + \sum_{k=1}^{m-k_1}\left(\sum_{p=0}^{k_1} a_p a_{p+k}\right)s_{n+k} + \sum_{k=m-k_1+1}^{m}\left(\sum_{p=0}^{m-k} a_p a_{p+k}\right)s_{n+k}\right). \end{aligned} \tag{46}$$

Again, only m data s_{n+1} to s_{n+m} on the right side of s_n participate in the interpolation, and the data after them are not needed. Therefore, (46) is the solution for all the optimal linear interpolations $\hat{s}_n = \hat{E}\{s_n \mid s_{n-i},\ -k_2 \le i \le k_1,\ i \neq 0\}$, where k_2 can be any integer greater than or equal to m.

The validity of (46) can also be proved as follows. The error is calculated as $e_n = s_n - \hat{s}_n$, where $\hat{s}_n$ is from (46). By extending the innovation noises from (13), it can be verified that

$$e_n = s_n - \hat{s}_n = \frac{1}{\sum_{k=0}^{k_1} a_k^2}\left(I'_n + a_1 I'_{n-1} + \cdots + a_{k_1} I'_{n-k_1}\right). \tag{47}$$

Using (15), it is quite clear that I'_{n-k_1} to I'_n are orthogonal to s_{n+k} for $k > 0$, and so is e_n in (47). Further, we have

$$\begin{aligned} &E\left\{s_{n-i}\left(I'_n + a_1 I'_{n-1} + \cdots + a_{k_1} I'_{n-k_1}\right)\right\} \\ &\quad = R_{I's}[i] + a_1 R_{I's}[i-1] + \cdots + a_{k_1} R_{I's}[i-k_1] \\ &\quad = R_{I's}[i] + a_1 R_{I's}[i-1] + \cdots + a_i R_{I's}[0], \quad 1 \le i \le k_1. \end{aligned} \tag{48}$$

The last equation of (48) is justified as we have $R_{I's}[k] = 0$ for $k < 0$ from (32). Using (32), (38), (47), and (48), it is seen that $E\{e_n s_{n-i}\} = 0$ for $1 \le i \le k_1$. This completes the proof.

The MSE is the same as in (45).

6.3. Infinite Data on Both Sides. Now, we want to estimate s_n from all the data around it. We will see that only m data on each side are needed and as is expected, the data with the same distance from s_n participate with the same weight. We have

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n-i},\, i \neq 0\}\\ &= -\frac{1}{\sum_{k=0}^{m} a_k^2}\cdot\left(\sum_{k=1}^{m}\left(\sum_{p=0}^{m-k} a_p a_{p+k}\right)(s_{n-k}+s_{n+k})\right).\end{aligned} \tag{49}$$

This estimation can also be obtained by letting $k_1 = m$ in (40) or (46). Again, note that (49) is the optimal solution for the problems $\hat{s}_n = \hat{E}\{s_n \mid s_{n-i},\, i \neq 0,\, -k_1 \leq i \leq k_2\}$, where k_1 and k_2 can be any integer greater than or equal to m.

The validity of (49) can also be proved as follows. The error is calculated as $e_n = s_n - \hat{s}_n$, where $\hat{s}_n$ is from (49). By extending the innovation noises from (4), it can be verified that

$$e_n = s_n - \hat{s}_n = \frac{1}{\sum_{k=0}^{m} a_k^2}(I_n + a_1 I_{n+1} + \cdots + a_m I_{n+m}). \tag{50}$$

Using (6) it is quite clear that I_n to I_{n+m} are orthogonal to s_{n-k} for $k > 0$, and so is e_n in (50). Further, we have

$$\begin{aligned}&E\{s_{n+i}(I_n + a_1 I_{n+1} + \cdots + a_m I_{n+m})\}\\ &\quad= R_{sI}[i] + a_1 R_{sI}[i-1] + \cdots + a_m R_{sI}[i-m], \quad i > 0.\end{aligned} \tag{51}$$

Using (32), (36), (50), and (51), it is seen that $E\{e_n s_{n+i}\} = 0$ for $i > 0$. This completes the proof.

The MSE is

$$\begin{aligned}E\{e_n^2\} &= \frac{1}{\left(\sum_{k=0}^{m} a_k^2\right)^2}\cdot E\left\{(I_n + a_1 I_{n+1} + \cdots + a_m I_{n+m})^2\right\}\\ &= \frac{\sigma^2}{\sum_{k=0}^{m} a_k^2}.\end{aligned} \tag{52}$$

7. Prediction with Finite Data

Assume that only $m-1$ consecutive data s_{n-1} to s_{n-m+1} are available. We want to prove the following.

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n-k},\, 1 \leq k \leq m-1\}\\ &= -\frac{1}{1-a_m^2}\sum_{k=1}^{m-1}(a_k - a_m a_{m-k})s_{n-k}.\end{aligned} \tag{53}$$

The above estimation can be obtained as follows. Since s_{n-m} is not available we can estimate it from data s_{n-1} to s_{n-m+1}. The estimated value can be now used to predict s_n using (16).

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n-k},\, 1 \leq k \leq m-1\}\\ &= -\sum_{k=1}^{m-1} a_k s_{n-k} - a_m \hat{s}_{n-m}\\ &= -a_1 s_{n-1} - a_2 s_{n-2} - \cdots - a_{m-1}s_{n-m+1} - a_m \hat{s}_{n-m}.\end{aligned} \tag{54}$$

On the other hand, s_{n-m} can be backward predicted using (30) as

$$\begin{aligned}\hat{s}_{n-m} &= \hat{E}\{s_{n-m} \mid s_{n-k},\, 1 \leq k \leq m-1\}\\ &= -a_1 s_{n-m+1} - a_2 s_{n-m+2} - \cdots - a_{m-1}s_{n-1} - a_m \hat{s}_n.\end{aligned} \tag{55}$$

Now we have two equations (54) and (55) with two unknowns $\hat{s}_n$ and $\hat{s}_{n-m}$. Solving these equations, we get (53). The optimality of (53) can also be proved by seeing that the estimation error is equal to

$$e_n = \frac{I_n - a_m I'_{n-m}}{1-a_m^2}. \tag{56}$$

To derive the above equation, we has used (4) and (13). It is easily seen from (6) and (15) that I_n and I'_{n-m} are orthogonal to data s_{n-1} to s_{n-m+1}. This proves the optimality of (53). To calculate the MSE, we note that

$$E\{e_n^2\} = E\{e_n(s_n - \hat{s}_n)\} = E\{e_n s_n\}. \tag{57}$$

The last equation is justified, as the error is orthogonal to the data and to the estimation which is a linear combination of the data. Inserting (56) in (57), we get

$$\begin{aligned}E\{e_n s_n\} &= \frac{1}{1-a_m^2}\cdot E\{(I_n - a_m I'_{n-m})s_n\}\\ &= \frac{1}{1-a_m^2}\cdot(R_{sI}[0] - a_m R_{I's}[-m]).\end{aligned} \tag{58}$$

Finally, using (58), (32), and (36), we have

$$E\{e_n^2\} = \frac{\sigma^2}{1-a_m^2}. \tag{59}$$

Higher-order predictions with $m-1$ data can be obtained from (53). As an example, we have

$$\begin{aligned}\hat{s}_n &= \hat{E}\{s_n \mid s_{n-k},\, 2 \leq k \leq m\}\\ &= -a_1\hat{s}_{n-1} - \sum_{k=2}^{m} a_k s_{n-k},\end{aligned} \tag{60}$$

where $\hat{s}_{n-1}$ is derived by replacing n by $n-1$ in (53).

We could not derive a simple general form for the estimation with less than $m-1$ data.

8. Interpolation with Finite Data

We now derive the linear interpolation with less than m data on each side. More clearly we allege

$$\begin{aligned}
\hat{s}_n &= \hat{E}\{s_n \mid s_{n+k},\ -k_2 \le k \le k_1,\ k \ne 0\} \\
&= -\frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2} \\
&\quad \times \left(\sum_{k=m-k_2}^{k_1} \left(\sum_{p=0}^{k_1-k} a_p a_{p+k} \right) s_{n+k} \right. \\
&\quad + \sum_{k=1}^{m-k_2-1} \left(\sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_2+1}^{m-k} a_p a_{p+k} \right) s_{n+k} \\
&\quad + \sum_{k=1}^{m-k_1-1} \left(\sum_{p=0}^{k_1} a_p a_{p+k} - \sum_{p=k_2-k+1}^{m-k} a_p a_{p+k} \right) s_{n-k} \\
&\quad \left. + \sum_{k=m-k_1}^{k_2} \left(\sum_{p=0}^{k_2-k} a_p a_{p+k} \right) s_{n-k} \right).
\end{aligned} \tag{61}$$

In (61) we must have $k_1 + k_2 \ge m - 1$ and $k_1 \le k_2 \le m - 1$. It means that the distance between s_n and the farthest data on the right side is less than the distance between s_n and the farthest data on the left side. The optimality of (61) can be seen as we can verify that from (61), (4), and (13) the estimation error is

$$\begin{aligned}
e_n &= \frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2} \\
&\quad \cdot \left(I_n + a_1 I_{n+1} + \cdots + a_{k_1} I_{n+k_1} - a_m I'_{n-m} \right. \\
&\quad \left. - a_{m-1} I'_{n-m+1} - \cdots - a_{k_2+1} I'_{n-k_2-1} \right).
\end{aligned} \tag{62}$$

It remains to prove that (62) is orthogonal to the data.

(1) It is clear from (6) and (15) that I_n to I_{n+k_1} and I'_{n-m} to I'_{n-k_2-1} are orthogonal to the data s_{n-1} to s_{n-k_2}. Therefore the error in (62) is orthogonal to s_{n-k} for $1 \le k \le k_2$.

(2) Further from (43) and regarding that I'_{n-m} to I'_{n-k_2-1} are orthogonal to the data s_{n+1} to s_{n+k_1} according to (15), we see that the error in (62) is orthogonal to s_{n+k} for $1 \le k \le k_1$.

Therefore the error is orthogonal to the data and the proof is completed.

From (32), (36), and (62), the MSE is

$$\begin{aligned}
E\{e_n^2\} &= E\{e_n s_n\} = \frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2} \\
&\quad \cdot \left(R_{sI}[0] + a_1 R_{sI}[-1] + \cdots + R_{sI}[-k_1] \right. \\
&\quad - a_m R_{I's}[-m] - a_{m-1} R_{I's}[-m+1] \cdots \\
&\quad \left. - a_{k_2+1} R_{I's}[-k_2-1] \right) \\
&= \frac{\sigma^2}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2}.
\end{aligned} \tag{63}$$

For the case $k_1 = k_2$, $2k_1 \ge m - 1$, $k_1 \le m - 1$, we can replace k_2 by k_1 in (61) to achieve the following.

$$\begin{aligned}
\hat{s}_n &= \hat{E}\{s_n \mid s_{n+k},\ -k_1 \le k \le k_1,\ k \ne 0\} \\
&= -\frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2} \\
&\quad \cdot \left(\sum_{k=m-k_1}^{k_1} \left(\sum_{p=0}^{k_1-k} a_p a_{p+k} \right) (s_{n-k} + s_{n+k}) \right. \\
&\quad \left. + \sum_{k=1}^{m-k_1-1} \left(\sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_1+1}^{m-k} a_p a_{p+k} \right) (s_{n-k} + s_{n+k}) \right).
\end{aligned} \tag{64}$$

As expected, we see that the data with the same distance from s_n participate with the same weight.

Now, consider the case that the distance between s_n and the farthest data on the right side is more than the distance between s_n and the farthest data on the left side. It can be handled by the symmetry of the problem. More clearly, if we replace s_{n-k} by s_{n+k} and vice versa in (61), we get the following.

$$\begin{aligned}
\hat{s}_n &= \hat{E}\{s_n \mid s_{n-k},\ -k_2 \le k \le k_1,\ k \ne 0\} \\
&= -\frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2} \\
&\quad \times \left(\sum_{k=m-k_2}^{k_1} \left(\sum_{p=0}^{k_1-k} a_p a_{p+k} \right) s_{n-k} \right. \\
&\quad + \sum_{k=1}^{m-k_2-1} \left(\sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_2+1}^{m-k} a_p a_{p+k} \right) s_{n-k} \\
&\quad + \sum_{k=1}^{m-k_1-1} \left(\sum_{p=0}^{k_1} a_p a_{p+k} - \sum_{p=k_2-k+1}^{m-k} a_p a_{p+k} \right) s_{n+k} \\
&\quad \left. + \sum_{k=m-k_1}^{k_2} \left(\sum_{p=0}^{k_2-k} a_p a_{p+k} \right) s_{n+k} \right).
\end{aligned} \tag{65}$$

Again in (65), $k_1 \leq k_2 \leq m-1$ and $k_1+k_2 \geq m-1$. The estimation error in this case is

$$e_n = \frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2} \cdot \left(I'_n + a_1 I'_{n-1} + \cdots + a_{k_1} I'_{n-k_1} - a_m I_{n+m} - a_{m-1} I_{n+m-1} - \cdots - a_{k_2+1} I_{n+k_2+1}\right). \quad (66)$$

The MSE is the same as (63). We could not find a simple general form for the case $k_1 + k_2 < m-1$.

9. A Detailed Example

In this section we deal with a detailed example. The optimal linear estimation of the following process is desired.

$$s_n + 0.8 s_{n-1} + 0.3 s_{n-2} - 0.1 s_{n-3} = I_n. \quad (67)$$

I_n is the innovation noise with the unit variance $\sigma = 1$. We have $a_1 = 0.8$, $a_2 = 0.3$ and $a_3 = -0.1$. The process is the response of the following 3rd order ($m = 3$) all-pole filter to the innovation noise.

$$H(z) = \frac{1}{1 + 0.8z^{-1} + 0.3z^{-2} - 0.1z^{-3}}. \quad (68)$$

The poles of this system are $p_1 = 0.2$ and $p_{2,3} = -0.5 \pm j0.5$. Taking inverse Z-transform from $S_s(z) = H(z)H(z^{-1})$, we get the following autocorrelation function.

$$\begin{aligned} R_s[k] &= r_k = E\{s_n s_{n-k}\} \\ &= \frac{625}{13542} \times 5^{-|n|} + \frac{40}{2257} \\ &\quad \times 2^{-|n|/2}\left(103\cos\left(\frac{3\pi}{4}n\right) - 26\sin\left(\frac{3\pi}{4}|n|\right)\right). \end{aligned} \quad (69)$$

From (69), we have $r_0 = 1.8716$, $r_1 = -1.1339$, $r_2 = 0.2322$, $r_3 = 0.3415$, $r_4 = -0.4563$, $r_5 = 0.2858$, and $r_6 = -0.0576$. Now, we consider different cases.

9.1. Prediction with Finite Data. We want to derive the following optimal linear prediction.

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n-1}, s_{n-2}\} = A_1 s_{n-1} + A_2 s_{n-2}. \quad (70)$$

Using (53), we have

$$\begin{aligned} \hat{s}_n &= -\frac{1}{1-0.01}\left[(0.8 + 0.1 \times 0.3)s_{n-1} + (0.3 + 0.1 \times 0.8)s_{n-2}\right] \\ &= -0.8384 s_{n-1} - 0.3838 s_{n-2}. \end{aligned} \quad (71)$$

If we want to verify the solution using the orthogonality equations, we have

$$E\{(s_n - A_1 s_{n-1} - A_2 s_{n-2}) s_{n-k}\} = 0, \quad k = 1, 2. \quad (72)$$

Expanding (72), we get

$$\begin{aligned} r_0 A_1 + r_1 A_2 &= r_1, \\ r_1 A_1 + r_0 A_2 &= r_2, \end{aligned} \quad (73)$$

where r_k's come from (69). Replacing r_k's from (69) in (73), we get

$$\begin{aligned} 1.8716 A_1 - 1.1339 A_2 &= -1.1339, \\ -1.1339 A_1 + 1.8716 A_2 &= 0.2322, \end{aligned} \quad (74)$$

Solving (74), we get the same result as (71).

9.2. Interpolation with Finite Data. Consider the following problem.

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n-1}, s_{n+1}\} = A_1 s_{n-1} + A'_1 s_{n+1} \quad (75)$$

It is the symmetric case of $k_1 = k_2 = 1$ and we have $2k_1 = 2 = m-1$. Using (64), we have

$$\begin{aligned} \hat{s}_n &= -\frac{[1 \times 0.8 - 0.3 \times (-0.1)]}{1 + 0.64 - 0.09 - 0.01}(s_{n-1} + s_{n+1}) \\ &= -0.5390(s_{n-1} + s_{n+1}). \end{aligned} \quad (76)$$

Let us rederive the solution of (75) using the orthogonality conditions. We have

$$E\{(s_n - A_1 s_{n-1} - A'_1 s_{n+1}) s_{n-k}\} = 0, \quad k = 1, -1. \quad (77)$$

Expanding (77), we get the following.

$$\begin{aligned} r_0 A_1 + r_2 A'_1 &= r_1, \\ r_2 A_1 + r_0 A'_1 &= r_1. \end{aligned} \quad (78)$$

Solving (78), we get the same answer as (76).

Now, consider the nonsymmetric following problem.

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n-2}, s_{n-1}, s_{n+1}\} = A_1 s_{n+1} + A'_1 s_{n-1} + A'_2 s_{n-2} \quad (79)$$

which is the case of $k_1 = 1 < k_2 = 2 \leq m-1$, and $k_1 + k_2 \geq m-1$. From (61), we get the following results.

$$\begin{aligned} \hat{s}_n &= -\frac{1}{1 + 0.64 - 0.01} \\ &\quad \cdot (1 \times 0.8 s_{n+1} + (1 \times 0.8 + 0.8 \times 0.3 - 0.3 \times (-0.1)) \\ &\quad \times s_{n-1} + 1 \times 0.3 s_{n-2}) \\ &= -0.4908 s_{n+1} - 0.6564 s_{n-1} - 0.1840 s_{n-2}. \end{aligned} \quad (80)$$

Now, we want to obtain the solution of (79) using the matrix equations and we expect the same answer as (80). The orthogonality condition is

$$E\{(s_n - A_1 s_{n+1} - A'_1 s_{n-1} - A'_2 s_{n-2}) s_{n-k}\} = 0, \quad k = -1, 1, 2. \quad (81)$$

It follows that

$$\begin{aligned} r_0 A_1 + r_2 A_1' + r_3 A_2' &= r_1, \\ r_2 A_1 + r_0 A_1' + r_1 A_2' &= r_1, \\ r_3 A_1 + r_1 A_1' + r_0 A_2' &= r_2. \end{aligned} \tag{82}$$

The result of (82) is the same as (80).

9.3. Interpolation with Infinite Data on the Left Side. We want to obtain the following estimation.

$$\begin{aligned} \hat{s}_n &= \hat{E}\{s_n \mid s_{n+i},\ i \le 1,\ i \ne 0\} \\ &= A_1 s_{n+1} + A_1' s_{n-1} + A_2' s_{n-2} + A_3' s_{n-3}. \end{aligned} \tag{83}$$

We can do it if we let $k_1 = 1$ in (40). It follows that

$$\begin{aligned} \hat{s}_n &= -\frac{1}{1+0.64} \\ &\quad \cdot (1 \times 0.8 s_{n+1} + (1 \times 0.8 + 0.8 \times 0.3) s_{n-1} \\ &\quad + (1 \times 0.3 + 0.8 \times (-0.1)) s_{n-2} + 1 \times (-0.1) s_{n-3}) \\ &= -0.4878 s_{n+1} - 0.6341 s_{n-1} - 0.1341 s_{n-2} + 0.0610 s_{n-3}. \end{aligned} \tag{84}$$

Now we verity (84) using the orthogonality conditions.

$$E\{(s_n - A_1 s_{n+1} - A_1' s_{n-1} - A_2' s_{n-2} - A_3' s_{n-3}) s_{n-k}\} = 0, \quad k = -1, 1, 2, 3. \tag{85}$$

The following set of equations is obtained

$$\begin{aligned} r_0 A_1 + r_2 A_1' + r_3 A_2' + r_4 A_3' &= r_1, \\ r_2 A_1 + r_0 A_1' + r_1 A_2' + r_2 A_3' &= r_1, \\ r_3 A_1 + r_1 A_1' + r_0 A_2' + r_1 A_3' &= r_2, \\ r_4 A_1 + r_2 A_1' + r_1 A_2' + r_0 A_3' &= r_3. \end{aligned} \tag{86}$$

Note that the coefficient matrix of (86) is not Toeplitz. The result of (86) is the same as (84).

10. Conclusion

We introduced anticausal LTI model besides the known causal LTI model for AR processes. Using these models and the related innovation noises, we achieved the optimal linear interpolations for different cases. Specifically, we extracted the formulae when there are infinite data on the right, or the left sides of the variable to be estimated. We also obtained the linear prediction or interpolation with finite data. The number of data must be at least the order of the process minus one. We could not find a general simple form when fewer data are available. For the proofs of our solutions, the innovation noises and the orthogonality principle are essential.

References

[1] Z.-D. Chen, R.-F. Chang, and W.-J. Kuo, "Adaptive predictive multiplicative autoregressive model for medical image compression," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 181–184, 1999.

[2] Z. Zhu and H. Leung, "Adaptive blind equalization for chaotic communication systems using extended-Kalman filter," *IEEE Transactions on Circuits and Systems I*, vol. 48, no. 8, pp. 979–989, 2001.

[3] D. Matrouf and J. L. Gauvain, "Using AR HMM state-dependent filtering for speech enhancement," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, pp. 785–788, March 1999.

[4] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 317–330, 1986.

[5] R. L. Burr and M. J. Cowan, "Autoregressive spectral models of heart rate variability: practical issues," *Journal of Electrocardiology*, vol. 25, pp. 224–233, 1992.

[6] M. B. Sirvanci and S. S. Wolff, "Nonparametric detection with autoregressive data," *IEEE Transactions on Information Theory*, vol. 2, no. 6, pp. 725–731, 1976.

[7] H. E. Witzgall and J. S. Goldstein, "Detection performance of the reduced-rank linear predictor ROCKET," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1731–1738, 2003.

[8] A. Golaup and A. H. Aghvami, "Modelling of MPEG4 traffic at GoP level using autoregressive processes," in *Proceedings of the 56th Vehicular Technology Conference*, pp. 854–858, September 2002.

[9] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, no. 3, pp. 223–229, 2002.

[10] J. H. Kim, "Forecasting autoregressive time series with bias-corrected parameter estimators," *International Journal of Forecasting*, vol. 19, no. 3, pp. 493–502, 2003.

[11] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 2002.

[12] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *Journal of Mathematical Physics*, vol. 25, no. 4, pp. 261–278, 1974.

[13] P. M. T. Broersen, "Finite-sample bias in the Yule-Walker method of autoregressive estimation," in *Proceedings of IEEE International Instrumentation and Measurement Technology Conference*, pp. 342–347, May 2008.

[14] P. M. T. Broersen, "Finite-sample bias propagation in autoregressive estimation with the Yule-Walker method," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 5, pp. 1354–1360, 2009.

[15] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Finite-sample bias in the Yule-Walker method of autoregressive estimation," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, pp. 001 815–001 818, May 2008.

[16] Y. Xia and M. S. Kamel, "A generalized least absolute deviation method for parameter estimation of autoregressive signals," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 107–118, 2008.

[17] T. Hsiao, "Identification of time-varying autoregressive systems using maximum a Posteriori estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3497–3509, 2008.

[18] J. Songsiri, J. Dahl, and L. Vandenberghe, "Maximum-likelihood estimation of autoregressive models with conditional independence constraints," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1701–1704, April 2009.

[19] A. Mahmoudi and M. Karimi, "Parameter estimation of autoregressive signals from observations corrupted with colored noise," *Signal Processing*, vol. 90, no. 1, pp. 157–164, 2010.

[20] W. X. Zheng, "Autoregressive parameter estimation from noisy data," *IEEE Transactions on Circuits and Systems II*, vol. 47, no. 1, pp. 71–75, 2000.