Research Article

Linear Estimation of Stationary Autoregressive Processes

Reza Dianat¹ and Farokh Marvasti²

¹ Engineering Department, Persian Gulf University, Daveas, 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

Copyright © 2011 R. Dianat and F. Marvasti. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Consider a sequence of an $m$th-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 $y_1, y_2, \ldots, y_n$ is the following linear combination

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

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:

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

We also write the above condition as

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

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 exploits 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 $m$th 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 orthogonality equations. We will find the optimal answer can be provided. Mahmoudi and Karimi propose an LS-based method to estimate AR parameters from noisy data [19]. The method exploits 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].

2. Causal and Anticausal Models

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

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

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 $\sigma$ 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

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

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.

\[ I_{n+k} \perp s_n, \quad k > 0, \quad n \in \mathbb{Z}. \]  

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

\[ H(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k}. \]  

For the system defined by (4), we have

\[ H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + \cdots + a_m z^{-m}}. \]  

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]

\[ S_z(z) = S(z)H(z)H(z^{-1}) = \sigma^2 H(z)H(z^{-1}). \]
In the above equation $S_s(z)$ is the PSDF of $s_n$ and $S_t(z)$ is the PSDF of $H_n$.

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 $l_n'$, note that the PSDF of $I_n'$ by using Figure 2 and (9) is as follows:

$$S_t(z) = S_s(z)H^{-1}(z^{-1})H^{-1}(z) = a^2H(z)H(z^{-1})H^{-1}(z)H^{-1}(z) = a^2. \tag{10}$$

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

$$H'(z) = \frac{1}{1 + a_1z + a_2z^2 + \cdots + a_mz^m}. \tag{12}$$

Using the above equation and Figure 3, we get

$$s_n + a_1s_{n+1} + a_2s_{n+2} + \cdots + a_ms_{n+m} = I_n', \quad n \in \mathbb{Z}. \tag{13}$$

Also, note that

$$s_n = \sum_{i=-\infty}^{\infty} h_i' I_{n-i} = h_0I_n' + h_1' I_{n+1} + h_2' I_{n+2} + \cdots \tag{14}$$

$$= h_0I_n' + h_1I_{n+1} + \cdots = \sum_{i=0}^{\infty} h_iI_{n+i}.$$

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

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

### 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

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n-k}, k > 0\} = -\sum_{k=1}^{m} a_k s_{n-k} \tag{16}$$

$$= -a_1s_{n-1} - a_2s_{n-2} - \cdots - a_m s_{n-m}.$$  

It can be 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]

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

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).

$$\hat{s}_{n-1} = \hat{E}\{s_{n-1} \mid s_{n-k}, k \geq 2\} = -a_1\hat{s}_{n-2} - \sum_{k=1}^{m} a_k s_{n-k-1} \tag{18}$$

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

$$e_n = s_n + \sum_{i=2}^{m} a_i s_{n-i} - a_1\sum_{k=1}^{m} a_k s_{n-k-1} = I_n - a_1I_{n-1}. \tag{19}$$

From (6), it is clear that $I_n$ and $I_{n-1}$ are orthogonal to $s_{n-k}$ for $k \geq 2$. 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.

$$\hat{s}_n = \hat{E}\{s_n \mid s_{n-k}, k \geq 3\}$$

$$= -a_1\hat{s}_{n-1} - a_2\hat{s}_{n-2} - \sum_{k=3}^{m} a_k s_{n-k} \tag{20}$$

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

$$e_n = s_n + a_1\hat{s}_{n-1} + a_2\hat{s}_{n-2} + \sum_{i=3}^{m} a_i s_{n-i}$$

$$= I_n - a_1I_{n-1} - a_2I_{n-2} + a_1\hat{s}_{n-1} + a_2\hat{s}_{n-2} \tag{21}$$

$$= I_n - a_1(s_{n-1} - \hat{s}_{n-1}) - a_2(s_{n-2} - \hat{s}_{n-2})$$

$$= I_n - a_1(I_{n-1} - a_1I_{n-2}) - a_2I_{n-2}.$$
4. The Problem Symmetries

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

\[
\hat{s}_n = \hat{E}[s_n | s_{n-k}, s_{n-k+1}, \ldots, s_{n-1}, s_n, \ldots, s_{n+k}]
\]

\[
= a'_{-k}s_{n-k} + a'_{-k+1}s_{n-k+1} + \cdots + a'_{k}s_{n+k}.
\]

The orthogonality principle gives

\[
E\{s_n - a'_{-k}s_{n-k} - a'_{-k+1}s_{n-k+1} - \cdots - a'_{k}s_{n+k}\} = 0, \quad i = -k_1, -k_1+1, \ldots, k_2, \quad i \neq 0.
\]

The above equations become

\[
R_i[i + k_1]a''_{-k_1} + R_i[i + k_1 - 1]a''_{-k_1+1} + \cdots + R_i[i - k_2]a''_{k_2} = R_i[i], \quad i = -k_1, -k_1+1, \ldots, k_2, \quad i \neq 0.
\]

Now, consider the following estimation.

\[
\hat{s}_n = \hat{E}[s_n | s_{n-k}, s_{n-k+1}, \ldots, s_{n-1}, s_n, \ldots, s_{n+k}]
\]

\[
= a''_{k_1}s_{n+k_1} + a''_{k_1-1}s_{n+k_1-1} + \cdots + a''_{k_2}s_{n-k_2}.
\]

The orthogonality of error to the data gives

\[
E\{s_n - a''_{k_1}s_{n+k_1} - a''_{k_1-1}s_{n+k_1-1} - \cdots - a''_{k_2}s_{n-k_2}\} = 0, \quad i = k_1, k_1-1, \ldots, -k_2, \quad i \neq 0.
\]

They become

\[
R_i[i - k_1]a''_{k_1} + R_i[i - k_1 + 1]a''_{k_1-1} + \cdots + R_i[i + k_2]a''_{k_2} = R_i[i], \quad i = k_1, k_1-1, \ldots, -k_2, \quad i \neq 0.
\]

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

\[
a'_{-k_1} = a''_{k_1}, a'_{-k_1+1} = a''_{k_1-1}, \ldots, a'_{k_2} = a''_{k_2}.
\]


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 \( R_{is}[k] = E\{s_nI_{n-k}\} \) and \( R_{Is}[k] = E\{I'_n s_{n-k}\} \). The first simple property follows from (16) and (15) as follows.

\[
R_{is}[k] = R_{Is}[k] = 0, \quad k < 0.
\]

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]

\[
R_{is}[k] = R_{Is}[k] = h[k] = \sigma^2 \delta[k] \ast h[k] = \sigma^2 h[k].
\]

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

\[
S_{is}(z) = \sigma^2 H(z) = \frac{\sigma^2}{1 + a_1 z^{-1} + \cdots + a_m z^{-m}}.
\]

Or equivalently

\[
S_{is}(z) (1 + a_1 z^{-1} + \cdots + a_m z^{-m}) = \sigma^2.
\]

Taking inverse Z-transform from this equation, we have

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

The right side of (36) is zero for \( k \neq 0 \). Referring to Figure 3, we have [11]

\[
R_{Is}[k] = R_{Is}[k] \ast h'[-k] = \sigma^2 \delta[k] \ast h'[-k] = \sigma^2 h'[-k].
\]

Again, we conclude that

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

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.

\[
\hat{s}_n = \hat{E}[s_n | s_{n+i}, i \leq k_1, i \neq 0].
\]

\[
\hat{s}_n = \hat{E}[s_n | s_{n+k}, k > 0] = -\sum_{k=1}^{m} a_{k} s_{n+k},
\]

\[
s_{n} + a_1 s_{n+1} + a_2 s_{n+2} + \cdots + a_m s_{n+m} = I'_n.
\]

Using (15), it is clear that the error is orthogonal to the data. It proves the optimality of (30).
\(k_1\) 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:

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

(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 \(\hat{s}_n = \hat{E}\{s_n \mid s_{n+i}, -k_2 \leq i \leq k_1, i \neq 0\}\), 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 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 \(I_k\) to \(I_{n+k}\) are orthogonal to \(s_{n-k}\) for \(k > 0\), and so is \(e_n\) in (41). Further, we have

\[
E\{s_{n+i}(I_n + a_1 I_{n+1} + \cdots + a_k I_{n+k})\} = R_d[i] + a_1 R_d[i-1] + \cdots + a_k R_d[i-k_1],
\]

\[
= R_d[i] + a_1 R_d[i-1] + \cdots + a_i R_d[0], \quad 1 \leq i \leq k_1.
\]

(42)

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

\[
E\{(I_n + a_1 I_{n+1} + \cdots + a_k I_{n+k}) s_{n+i}\} = 0, \quad 1 \leq i \leq k_1.
\]

(43)

This completes the proof.

The MSE is

\[
E\{e_n^2\} = \frac{1}{\sum_{k=0}^{k_1} a_k^2} \cdot E\{\{(I_n + a_1 I_{n+1} + \cdots + a_k I_{n+k})\}^2\}
\]

\[
= \frac{1}{\sum_{k=0}^{k_1} a_k^2} \cdot \left( \sum_{k=1}^{k_1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} \right)^2 \right) \sum_{k=1}^{m-k} \left( \sum_{p=0}^{k} a_p a_{p+k} \right)^2
\]

\[\cdot \left( \sum_{k=1}^{m} \left( \sum_{p=0}^{k-m-k+1} a_p a_{p+k} \right)^2 \right).
\]

(44)

Therefore,

\[
E\{e_n^2\} = \frac{\sigma^2}{\sum_{k=0}^{k_1} a_k^2}.
\]

(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.

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

(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 \(\hat{s}_n = \hat{E}\{s_n \mid s_{n-i}, -k_2 \leq i \leq k_1, i \neq 0\}\), 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

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

(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

\[
E\{s_{n-i}\left(I_n' + a_1 I_{n-1} + \cdots + a_k I_{n-k} \right)\} = R_{d'}[i] + a_1 R_{d'}[i-1] + \cdots + a_k R_{d'}[i-k_1],
\]

\[
= R_{d'}[i] + a_1 R_{d'}[i-1] + \cdots + a_i R_{d'}[0], \quad 1 \leq i \leq k_1.
\]

(48)

The last equation of (48) is justified as we have \(R_{d'}[k] = 0\) for \(k < 0\) from (32). Using (32), (38), (47), and (48), it is seen that

\[
E\{e_{n_{s_n-i}}\} = 0 \quad \text{for} \quad 1 \leq i \leq k_1.
\]

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

$$\hat{s}_n = \hat{E} \{ s_n \mid s_{n-i}, i \neq 0 \}$$

$$= \frac{1}{\sum_{k=0}^{m} a_k} \cdot \left( \sum_{k=1}^{m-k} \left( \sum_{p=0}^{k} a_p a_{p+k} \right) (s_{n-k} + s_{n+k}) \right).$$

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 problem $\hat{s}_n = \hat{E} \{ s_n \mid s_{n-i}, i \neq 0, -k_1 \leq i \leq k_1 \}$, where $k_1$ and $k_2$ can be any integer greater than or equal to $m$.

The validation 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

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

(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

$$E \{ s_{n+i} \mid I_n + a_1 I_{n+1} + \cdots + a_m I_{n+m} \}$$

$$= R_{I,i} [i] + a_1 R_{I,i} [i-1] + \cdots + a_m R_{I,i} [i-m], \quad i > 0.$$ (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

$$E \{ e_n^2 \} = \frac{1}{\sum_{k=0}^{m} a_k^2} \cdot \left( \sum_{k=0}^{m} a_k^2 \right).$$

(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.

$$\hat{s}_n = \hat{E} \{ s_n \mid s_{n-k}, 1 \leq k \leq m - 1 \}$$

$$= \frac{1}{1 - a_m^2} \sum_{k=1}^{m-1} (a_k - a_m a_{m-k}) s_{n-k}.$$ (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 now be used to predict $s_n$ using (16).

$$\hat{s}_n = \hat{E} \{ s_n \mid s_{n-k}, 1 \leq k \leq m - 1 \}$$

$$= \frac{1}{1 - a_m^2} \sum_{k=1}^{m-1} a_k s_{n-k} - a_m \hat{s}_{n-m}$$

$$= -a_1 \hat{s}_{n-1} - a_2 \hat{s}_{n-2} - \cdots - a_{m-1} \hat{s}_{n-m+1} - a_m \hat{s}_{n-m}.$$ (54)

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

$$\hat{s}_{n-m} = \hat{E} \{ s_{n-m} \mid s_{n-k}, 1 \leq k \leq m - 1 \}$$

$$= -a_1 s_{n-m} - a_2 s_{n-m+1} - \cdots - a_{m-1} s_{n-m+1} - a_m \hat{s}_n.$$ (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

$$e_n = \frac{I_n - a_m I_{n-m}}{1 - a_m^2}.$$ (56)

To derive the above equation, we have 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

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

$$E \{ e_n s_n \} = \frac{1}{1 - a_m^2} \cdot E \{ (I_n - a_m I_{n-m}) s_n \}$$

$$= \frac{1}{1 - a_m^2} \cdot (R_{I,0} [0] - a_m R_{I,-1} [-m]).$$ (58)

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

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

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

$$\hat{s}_n = \hat{E} \{ s_n \mid s_{n-k}, 2 \leq k \leq m \}$$

$$= -a_1 \hat{s}_{n-1} - \sum_{k=2}^{m} a_k \hat{s}_{n-k},$$ (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

\[ \hat{s}_n = \hat{E}\{s_n \mid s_{n+k}, -k_2 \leq k \leq k_1, k \neq 0\} \]

\[ = -\frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2} \times \left( \sum_{k=m-k_2}^{k-1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_2+1}^{m-k} a_p a_{p+k} \right) s_{n+k} \right. \]

\[ \left. + \sum_{k=1}^{m-k_1-1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_2-k}^{m-k} a_p a_{p+k} \right) s_{n-k} \right) \cdot \left( \sum_{k=m-k_1}^{k_1-k} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} \right) \right) \]

\[ + \sum_{k=m-k_1}^{k_1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} \right) s_{n-k} \] \hspace{1cm} \text{(61)}

In (61) we must have \( k_1 + k_2 \geq m - 1 \) and \( k_1 \leq k_2 \leq m - 1 \). 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

\[ e_n = \frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2} \cdot \left( I_n + a_1 I_{n+1} + \cdots + a_{k_1} I_{n+k_1} - a_{m-1} I_{n-m-1} \right. \]

\[ \left. - a_{m-1} I_{n-m-1} - \cdots - a_{k_2} I_{n-k_2-1} \right) \] \hspace{1cm} \text{(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 \( 1 \leq k \leq k_2 \).

(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 \( 1 \leq k \leq k_1 \).

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

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

\[ E[e_n^2] = E[e_n s_n] = \frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2} \]

\[ \cdot \left( R_0^s[0] + a_1 R_0^s[-1] + \cdots + R_0^s[-k_1] \right. \]

\[ \left. - a_{m-1} R_0^s[-m] - a_{m-1} R_0^s[-m+1] \right) \cdots \hspace{1cm} \text{(63)} \]

\[ - a_{k_2+1} R_0^s[-k_2] \]

\[ = \frac{\sigma^2}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2}. \]

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

\[ \hat{s}_n = \hat{E}\{s_n \mid s_{n+k}, -k_1 \leq k \leq k_1, k \neq 0\} \]

\[ = -\frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2} \times \left( \sum_{k=m-k_1}^{k_1-k} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} \right) (s_{n-k} + s_{n+k}) \right. \]

\[ \left. + \sum_{k=1}^{m-k_1-1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_2-k}^{m-k} a_p a_{p+k} \right) (s_{n-k} + s_{n+k}) \right) \] \hspace{1cm} \text{(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.

\[ \hat{s}_n = \hat{E}\{s_n \mid s_{n+k}, -k_2 \leq k \leq k_1, k \neq 0\} \]

\[ = -\frac{1}{\sum_{k=0}^{k_1} a_k^2 - \sum_{k=k_1+1}^{m} a_k^2} \times \left( \sum_{k=m-k_2}^{k-1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} \right) s_{n-k} \right. \]

\[ \left. + \sum_{k=1}^{m-k_1-1} \left( \sum_{p=0}^{k_1-k} a_p a_{p+k} - \sum_{p=k_2-k}^{m-k} a_p a_{p+k} \right) s_{n-k} \right) \] \hspace{1cm} \text{(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
\[
e_n = \frac{1}{\sum_{k=0}^{k_2} a_k^2 - \sum_{k=k_2+1}^{m} a_k^2} \cdot \left( I_n' + a_1 I_{n-1}' + \cdots + a_{k_1} I_{k_1}' \right.
\]
\[\left. - a_m I_{n+m} - a_{m-1} I_{n+m-1} - \cdots - a_{k_2+1} I_{n+k_2+1} \right) . \tag{66}\]

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.
\[
s_n + 0.8s_{n-1} + 0.3s_{n-2} - 0.1s_{n-3} = I_n. \tag{67}\]

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

The poles of this system are $p_1 = 0.2$ and $p_{2,3} = -0.5 \pm 0.5i$. Taking inverse $Z$-transform from $S(z) = H(z)H(z^{-1})$, we get the following autocorrelation function.
\[
R_i[k] = r_k = E[s_n s_{n-k}]
\]
\[= \frac{625 \times 5^{-|n|} + 40}{2257} \times 2^{-|n|/2} \left( 103 \cos \left( \frac{3\pi}{4} |n| \right) - 26 \sin \left( \frac{3\pi}{4} |n| \right) \right) . \tag{69}\]

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.
\[
\hat{s}_n = \hat{E}[s_n \mid s_{n-1}, s_{n-2}] = A_1 s_{n-1} + A_2 s_{n-2}. \tag{70}\]

Using (53), we have
\[
\hat{s}_n = -\frac{1}{1 - 0.01} \left[ (0.8 + 0.1 \times 0.3) s_{n-1} + (0.3 + 0.1 \times 0.8) s_{n-2} \right]
\]
\[= -0.8384 s_{n-1} - 0.3838 s_{n-2}. \tag{71}\]

If we want to verify the solution using the orthogonality equations, we have
\[
E\{ (s_n - A_1 s_{n-1} - A_2 s_{n-2}) s_{n-k} \} = 0, \quad k = 1, 2. \tag{72}\]

Expanding (72), we get
\[
r_0 A_1 + r_1 A_2 = r_1, \tag{73}\]
\[
r_1 A_1 + r_2 A_2 = r_2, \tag{74}\]
where $r_k$'s come from (69). Replacing $r_k$'s from (69) in (73), we get
\[
1.8716 A_1 - 1.1339 A_2 = -1.1339, \tag{75}\]
\[-1.1339 A_1 + 1.8716 A_2 = 0.2322, \tag{76}\]
Solving (74), we get the same result as (71).

9.2. Interpolation with Finite Data. Consider the following problem.
\[
\hat{s}_n = \hat{E}[s_n \mid s_{n-1}, s_{n+1}] = A_1 s_{n-1} + A_2 s_{n+1} \tag{77}\]

It is the symmetric case of $k_1 = k_2 = 1$ and we have $2k_1 = 2 = m - 1$. Using (64), we have
\[
\hat{s}_n = \frac{[1 \times 0.8 - 0.3 \times (-0.1)]}{1 + 0.64 - 0.09 - 0.01} (s_{n-1} + s_{n+1})
\]
\[= -0.5390 (s_{n-1} + s_{n+1}). \tag{78}\]

Let us derive the solution of (75) using the orthogonality conditions. We have
\[
E\{ (s_n - A_1 s_{n-1} - A_2 s_{n+1}) s_{n-k} \} = 0, \quad k = 1, -1. \tag{79}\]

Expanding (77), we get the following.
\[
r_0 A_1 + r_2 A_1^* = r_1, \tag{80}\]
\[
r_2 A_1 + r_0 A_1^* = r_1. \tag{81}\]

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

Now, consider the nonsymmetric following problem.
\[
\hat{s}_n = \hat{E}[s_n \mid s_{n-2}, s_{n-1}, s_{n+1}] = A_1 s_{n-1} + A_1^* s_{n-1} + A_2 s_{n-2} \tag{82}\]

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.
\[
\hat{s}_n = \frac{1}{1 + 0.64 - 0.01} \left[ (1 \times 0.8 s_{n+1} + (1 \times 0.8 + 0.8 \times 0.3 - 0.3 \times (-0.1)) \right.
\]
\[\times s_{n-1} + 1 \times 0.3 s_{n-2})
\]
\[= -0.4908 s_{n+1} - 0.6564 s_{n-1} - 0.1840 s_{n-2}. \tag{83}\]

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
\[
E\{ (s_n - A_1 s_{n+1} - A_2 s_{n-2}) s_{n-k} \} = 0, \quad k = -1, 1, 2. \tag{84}\]
It follows that

\[
\begin{align*}
    r_0 A_1 + r_2 A_1' + r_3 A_2' &= r_1, \\
    r_2 A_1 + r_0 A_1' + r_1 A_1' &= r_1, \\
    r_3 A_1 + r_1 A_1' + r_0 A_2' &= r_2.
\end{align*}
\] (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.

\[
\hat{s}_n = E\{s_n \mid s_{n+i}, i \leq 1, i \neq 0\}
\]

\[
= A_1 s_{n+1} + A_1' s_{n-1} + A_2 s_{n-2} + A_3 s_{n-3}.
\] (83)

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

\[
\hat{s}_n = -\frac{1}{1 + 0.64}
\]

\[
\cdot (1 \times 0.8 s_{n+1} + (1 \times 0.8 + 0.8 \times 0.3) s_{n-1})
\]

\[
+ (1 + 0.8 + 0.3 \times (-0.1)) s_{n-2} + 1 \times (-0.1) s_{n-3}
\]

\[
= -0.4878 s_{n+1} - 0.6341 s_{n-1} - 0.1341 s_{n-2} + 0.0610 s_{n-3}.
\] (84)

Now we verify (84) using the orthogonality conditions.

\[
E\{s_n - A_1 s_{n+1} - A_1' s_{n-1} - A_2 s_{n-2} - A_3 s_{n-3}\} = 0,
\]

\[
k = -1, 1, 2, 3.
\] (85)

The following set of equations is obtained

\[
\begin{align*}
    r_0 A_1 + r_2 A_1' + r_3 A_2' + r_4 A_3' &= r_1, \\
    r_2 A_1 + r_0 A_1' + r_1 A_1' + r_2 A_2' &= r_1, \\
    r_3 A_1 + r_1 A_1' + r_0 A_2' + r_1 A_3' &= r_2, \\
    r_4 A_1 + r_2 A_1' + r_1 A_2' + r_0 A_3' &= r_3.
\end{align*}
\] (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.

