Linear predictive analysis of speech using non-Toeplitz autocorrelation matrices

Mitsuo Komura

*International Institute for Advanced Study of Social Information Science,
Fujitsu Limited, 140, Miyamoto, Numazu, 410-03 Japan*

(Received 2 July 1985)

A covariance method is proposed to analyze speech more accurately than an autocorrelation method. The new covariance method, called the non-Toeplitz form method, ensures uniqueness and stability of data analysis. First, forward and backward prediction models are presented. The uniqueness of definitions of a data vector and a prediction error vector is proved. Second, prediction procedures based on Gram-Schmidt orthogonalization are discussed, and new sample autocorrelation functions are defined. Third, prediction procedures for solving normal equations by using matrix decomposition are presented. An equivalence relationship between auxiliary variables in this procedure and PARCOR coefficients is also discussed. Fourth, the positive definiteness of an autocorrelation matrix whose elements are given by the autocorrelation function proposed in this paper is shown. Finally, the proposed method is compared with other methods.

PACS number: 43. 60. Cg, 43. 60. Gk

1. INTRODUCTION

The autocorrelation method and the covariance method are the most used methods for linear prediction of speech. In the autocorrelation method, we assume that the input signal is stationary, and the sample autocorrelation function has only one variable, which depends only on time difference. The autocorrelation method can be executed efficiently by using Levinson's recurrence formula. The PARCOR method¹⁾ is also based on this recurrence formula. In the covariance method, the sample autocorrelation function is assumed to be a function of two variables. It can also be applied to input that is not stationary. However, the covariance methods proposed in recent studies have the following weak points:

- i) The covariance methods proposed by Atal and Hanauer,²⁾ Morf *et al.*³⁾ do not satisfy the equivalence relationship between least squares estimation, estimation based on the orthogonality

principle, and estimation based on the Gram-Schmidt procedure.

- ii) Predictor coefficients cannot be obtained uniquely, because the autocorrelation matrices used in the covariance methods proposed by others²⁻⁶⁾ are not positive definite.

- iii) Levinson's recurrence formula applied to stationary input is used in the covariance method proposed by Markel and Gray.⁴⁾

This paper proposes a new covariance method which has the following features:

- (a) It satisfies the equivalence relationship in i).
- (b) The autocorrelation matrices composed of the proposed autocorrelation functions are positive definite. Therefore we can always obtain unique solutions.
- (c) Levinson's recurrence formula is not used.

We shall call this new method the non-Toeplitz form method.

In the next section, the problem is stated, and new forward and backward prediction models are shown.

Unique definition of data vectors and prediction error vectors is also shown.

In Section 3, forward and backward prediction procedures based on orthogonalization are shown. New forward and backward sample autocorrelation functions are defined.

In Section 4, procedures for solving normal equations by using matrix decomposition are presented. The equivalence relationship between the PARCOR coefficients and the auxiliary variables used in the matrix decomposition procedure is discussed.

In Section 5, the positive definiteness of the autocorrelation matrices whose elements are given by the autocorrelation functions proposed in this paper is shown.

In Section 6, covariance methods based on orthogonalization and methods using normal equations are discussed. The covariance method proposed in this paper is compared with other methods.

2. PROBLEM STATEMENT AND UNIQUENESS OF THE DEFINITION OF MODELS

Let $X(t, \omega)$ be a stochastic process, where t is a time parameter, and ω is a probability parameter. For $X(t, \omega)$ we can consider four cases;

- (1) A family of time functions (t and ω variables):

$$X(t, \omega) \Rightarrow X(t),$$

- (2) A random variable (t fixed, ω variable):

$$X(t_n, \omega) \Rightarrow X(t_n) \text{ or } X(n),$$

- (3) A single time function (t variable, ω fixed):

$$X(t, \omega_n) \Rightarrow x(t),$$

- (4) A single number (t fixed, ω fixed):

$$X(t_n, \omega_n) \Rightarrow x(t_n) \text{ or } x(n).$$

In this paper, each continuous-time signal $x(t)$ is sampled to obtain a discrete-time signal $x(nT)$ ($t_n = nT$), where n is an integer and T is the sampling interval chosen to be $T=1$. We shall discuss linear prediction problems on this sample space.

2.1 Definition of Models

We now define an m -th-order forward predictor of a sample $x(n)$ by

$$\hat{x}_m(n) = - \sum_{i=1}^m a_{mi} x(n-i) \quad (m=1, 2, \dots, M), \quad (1)$$

where $-a_{mi}$ is the i -th forward predictor coefficient and M is the maximum order of the linear predictor.

Let $e_m^+(n)$ be the forward prediction error between the sample $x(n)$ and the predictor $\hat{x}_m(n)$. Then, $x(n)$ and $e_m^+(n)$ are given by

$$x(n) = \hat{x}_m(n) + e_m^+(n), \quad (2)$$

$$e_m^+(n) = \sum_{i=0}^m a_{mi} x(n-i), \quad a_{m0} = 1. \quad (3)$$

Similarly, we define an m -th-order backward prediction model by

$$\begin{aligned} \tilde{x}_m(n) &= - \sum_{i=1}^m b_{mi} x(n+m+1-i) \\ (m &= 1, 2, \dots, M), \end{aligned} \quad (4)$$

$$x(n) = \tilde{x}_m(n) + e_m^-(n), \quad (5)$$

$$e_m^-(n) = \sum_{i=1}^{m+1} b_{mi} x(n+m+1-i), \quad b_{m,m+1} = 1, \quad (6)$$

where $\tilde{x}_m(n)$ is a backward predictor, $e_m^-(n)$ is the backward prediction error, and $-b_{mi}$ is the i -th backward predictor coefficient.

Markel and Gray⁴⁾ *et al.* used a different definition for the backward prediction error $x_m^-(n)$

$$x_m^-(n) = x(n-m-1) - \sum_{i=1}^m b_{mi} x(n-i) \quad (7-a)$$

$$= \sum_{i=1}^{m+1} b_{mi} x(n-i), \quad b_{m,m+1} = 1, \quad (7-b)$$

where the first term of the right-hand side of Eq. (7-a) is the sample, and the second term is its predictor. It should be noted that there is a time difference of $m+1$ between prediction error $e_m^-(n)$ and prediction error $x_m^-(n)$. Thus, $e_m^-(n)$ and $x_m^-(n)$ are related as follows:

$$e_m^-(n) = x_m^-(n+m+1). \quad (8)$$

The first problem is to find a set of predictor coefficients that will minimize the sum of the squared prediction errors.

We consider four types of total squared errors and the corresponding normal equations, as follows. (Total squared errors and the normal equations proposed in this paper)

(I-f) For forward prediction,

$$S_m^+ = \sum_{n=0}^{N-1} \{e_m^+(n)\}^2. \quad (9)$$

$\partial S_m^+ / \partial a_{mi} = 0$; that is,

$$\sum_{i=0}^m a_{mi} r_{m-i, m-j} = 0 \quad (j=1, 2, \dots, m), \quad (10)$$

where

$$r_{i,j} = \frac{1}{N} \sum_{n=0}^{N-1} x(n+i)x(n+j) .$$

(I-b) For backward prediction,

$$S_m^- = \sum_{n=0}^{N-1} \{e_{\langle m \rangle}^-(n)\}^2 . \quad (11)$$

$\partial S_m^- / \partial b_{mi} = 0$; that is,

$$\sum_{i=1}^{m+1} b_{mi} r_{1-i,1-j} = 0 \quad (j=1, 2, \dots, m) , \quad (12)$$

where

$$r_{-i,-j} = \frac{1}{N} \sum_{n=0}^{N-1} x(n-i)x(n-j) .$$

$e_{\langle m \rangle}^+$ is defined by Eq. (18), that is, $e_{\langle m \rangle}^+(n) = e_m^+(n+m)$.

$e_{\langle m \rangle}^-$ is defined by Eq. (24), that is, $e_{\langle m \rangle}^-(n) = e_m^-(n-m)$.

(Total squared errors and the normal equations used by others)

(II-f) By Atal and Hanauer,²⁾ Chandra and Lin,⁵⁾

$$S_m'^+ = \sum_{n=0}^{N-1} \{e_m^+(n)\}^2 . \quad (13)$$

$\partial S_m'^+ / \partial a_{mi} = 0$; that is,

$$\sum_{i=0}^m a_{mi} r_{-i,-j} = 0 \quad (j=1, 2, \dots, m) . \quad (14)$$

(II-b) By Markel and Gray,⁴⁾

$$S_m'^- = \sum_{n=0}^{N-1} \{x_m^-(n)\}^2 . \quad (15)$$

$\partial S_m'^- / \partial b_{mi} = 0$; that is,

$$\sum_{i=1}^{m+1} b_{mi} r_{-i,-j} = 0 \quad (j=1, 2, \dots, m) . \quad (16)$$

Fig. 1 Ranges of data used to obtain sample autocorrelation functions. (a) range for type (I-f). (b) range for type (I-b). (c) range for type (II-f). (d) range for type (II-b). * $(\cdot, \cdot)$ denotes inner product. Vector $\mathbf{x}_{\langle m \rangle}$ and $\mathbf{x}_{\langle m \rangle}$ are defined by Eqs. (17) and (23) respectively.

The ranges of the data used to solve the normal equations are shown in Fig. 1. It is clear from Figs. 1 (a) and 1 (c) that the ranges for type (I-f) and type (II-f) do not match. It is also clear from Figs. 1 (b) and 1 (d) that there is a time shift between the ranges for type (I-b) and type (II-b). In Sections 2.2 and 2.3, we will show that we should use the normal equations of (I-f) and (I-b) proposed in this paper, instead of those of (II-f) and (II-b).

2.2 Uniqueness of Definition of Data Vectors and Prediction Error Vectors

In this section, we present a unique definition for data vectors and prediction error vectors to which we can apply the Gram-Schmidt procedure.

From Theorem 2.1 below, the forward data vector $\mathbf{x}_{\langle m \rangle}$ and the forward prediction error vector $\mathbf{e}_{\langle m \rangle}^+$ are represented as follows:*

$$\begin{aligned}\mathbf{x}_{\langle m \rangle} &= (x_{\langle m \rangle}(0), x_{\langle m \rangle}(1), \dots, x_{\langle m \rangle}(N-1))^T \\ &= (x(m), x(1+m), \dots, x(N-1+m))^T, \quad (17)\end{aligned}$$

$$\begin{aligned}\mathbf{e}_{\langle m \rangle}^+ &= (e_{\langle m \rangle}^+(0), e_{\langle m \rangle}^+(1), \dots, e_{\langle m \rangle}^+(N-1))^T \\ &= (e_m^+(m), e_m^+(1+m), \dots, e_m^+(N-1+m))^T, \quad (18)\end{aligned}$$

for $m=0, 1, \dots, M$. The ranges of the data vectors $\{\mathbf{x}_{\langle m \rangle}\}$ are shown in Fig. 1 (a).

Theorem 2.1

Let

$$\begin{aligned}\mathbf{x}_{\langle m \rangle} &= (x(f(m)), x(1+f(m)), \\ &\dots, x(N-1+f(m)))^T, \quad (19)\end{aligned}$$

$$\begin{aligned}\mathbf{e}_{\langle m \rangle}^+ &= (e_m^+(g(m)), e_m^+(1+g(m)), \\ &\dots, e_m^+(N-1+g(m)))^T, \quad (20)\end{aligned}$$

and

$$\mathbf{x}_{\langle 0 \rangle} = (x(0), x(1), \dots, x(N-1))^T. \quad (21)$$

Suppose that $\{\mathbf{e}_{\langle m \rangle}^+\}_{m=0}^M$ is obtained from $\{\mathbf{x}_{\langle m \rangle}\}_{m=0}^M$ using the Gram-Schmidt orthogonalization procedure:

$$\begin{aligned}\mathbf{e}_{\langle 0 \rangle}^+ &= \mathbf{x}_{\langle 0 \rangle}, \\ \mathbf{e}_{\langle m \rangle}^+ &= \mathbf{x}_{\langle m \rangle} - \sum_{l=0}^{m-1} \gamma_{lm} \mathbf{e}_{\langle l \rangle}^+, \quad m=1, 2, \dots, M,\end{aligned}$$

where

* To be exact, $\mathbf{x}_{\langle m \rangle}$ and $\mathbf{e}_{\langle m \rangle}^+$ are represented as $\mathbf{x}_{\langle m \rangle, t_0, T}$, and $\mathbf{e}_{\langle m \rangle, t_0, T}^+$ respectively. Without loss of generality, we can set $t_0=0$ and $T=N-1$, so we omit the subscripts t_0 and T .

$$\gamma_{lm} = \frac{(\mathbf{e}_{\langle l \rangle}^+, \mathbf{x}_{\langle m \rangle})}{(\mathbf{e}_{\langle l \rangle}^+, \mathbf{e}_{\langle l \rangle}^+)},$$

$(\cdot, \cdot)$ denotes an inner product. Then it follows that

$$f(m)=g(m)=m. \quad (22)$$

Proof. The proof is shown in Appendix of Ref. 7).
Q.E.D.

From Theorem 2.2 below, the backward data vector $\mathbf{x}_{\langle m \rangle}$ and the backward prediction error vector $\mathbf{e}_{\langle m \rangle}^-$ are represented as follows:

$$\begin{aligned}\mathbf{x}_{\langle m \rangle} &= (x_{\langle m \rangle}(0), x_{\langle m \rangle}(1), \dots, x_{\langle m \rangle}(N-1))^T \\ &= (x(-m), x(1-m), \dots, x(N-1-m))^T, \quad (23)\end{aligned}$$

$$\begin{aligned}\mathbf{e}_{\langle m \rangle}^- &= (e_{\langle m \rangle}^-(0), e_{\langle m \rangle}^-(1), \dots, e_{\langle m \rangle}^-(N-1))^T \\ &= (e_m^-(-m), e_m^-(1-m), \dots, e_m^-(N-1-m))^T, \quad (24)\end{aligned}$$

for $m=0, 1, \dots, M$. The ranges of the data vectors $\{\mathbf{x}_{\langle m \rangle}\}$ are shown in Fig. 1 (b).

The following theorem can be proved in the same way as Theorem 2.1.

Theorem 2.2

Let

$$\begin{aligned}\mathbf{x}_{\langle m \rangle} &= (x(h(m)), x(1+h(m)), \\ &\dots, x(N-1+h(m)))^T, \quad (25)\end{aligned}$$

$$\begin{aligned}\mathbf{e}_{\langle m \rangle}^- &= (e_m^-(k(m)), e_m^-(1+k(m)), \\ &\dots, e_m^-(N-1+k(m)))^T, \quad (26)\end{aligned}$$

and

$$\begin{aligned}\mathbf{x}_{\langle 0 \rangle} &= (x(0), x(1), \dots, x(N-1))^T; \\ \text{that is, } h(0) &= 0. \quad (27)\end{aligned}$$

Suppose that $\{\mathbf{e}_{\langle m \rangle}^-\}_{m=0}^M$ is obtained from $\{\mathbf{x}_{\langle m \rangle}\}_{m=0}^M$ using the Gram-Schmidt orthogonalization procedure:

$$\begin{aligned}\mathbf{e}_{\langle 0 \rangle}^- &= \mathbf{x}_{\langle 0 \rangle}, \\ \mathbf{e}_{\langle m \rangle}^- &= \mathbf{x}_{\langle m \rangle} - \sum_{l=0}^{m-1} \delta_{lm} \mathbf{e}_{\langle l \rangle}^-, \quad m=1, 2, \dots, M,\end{aligned}$$

where

$$\delta_{lm} = \frac{(\mathbf{e}_{\langle l \rangle}^-, \mathbf{x}_{\langle m \rangle})}{(\mathbf{e}_{\langle l \rangle}^-, \mathbf{e}_{\langle l \rangle}^-)}.$$

Then it follows that

$$h(m)=k(m)=-m. \quad (28)$$

In the next section the equivalence relationship between estimation based on the Gram-Schmidt

Table 1 Equivalence relationships between three estimation procedures.

	Type (I-f)	Type (I-b)
Least squares estimation	$\alpha_m = \frac{1}{N} \sum_{n=0}^{N-1} \{e_{\langle m}^+(n)\}^2$ $\partial \alpha_m / \partial a_{mi} = 0 \quad (29)$	$\beta_m = \frac{1}{N} \sum_{n=0}^{N-1} \{e_{\langle m}^-(n)\}^2$ $\partial \beta_m / \partial b_{mi} = 0 \quad (32)$
Estimation based on orthogonality principle	$\sum_{n=0}^{N-1} \left\{ \sum_{i=0}^m a_{mi} x(n+m-i) x(n+m-j) \right\} = 0$ $(j=1, 2, \dots, m) \quad (30)$	$\sum_{n=0}^{N-1} \left\{ \sum_{i=1}^{m+1} b_{mi} x(n+1-i) x(n+1-j) \right\} = 0$ $(j=1, 2, \dots, m) \quad (33)$
Estimation based on Gram-Schmidt procedure	$\sum_{n=0}^{N-1} e_{\langle m}^+(n) e_{\langle l}^+(n) = 0$ $(l \neq m) \quad (31)$	$\sum_{n=0}^{N-1} e_{\langle m}^-(n) e_{\langle l}^-(n) = 0$ $(l \neq m) \quad (34)$
Data vector	$\mathbf{x}_{\langle m} = (x_{\langle m}(0), \dots, x_{\langle m}(N-1))^T$ $= (x(m), \dots, x(N-1+m))^T$	$\mathbf{x}_{\langle m} = (x_{\langle m}(0), \dots, x_{\langle m}(N-1))^T$ $= (x(-m), \dots, x(N-1-m))^T$
Prediction error vector	$\mathbf{e}_{\langle m}^+ = (e_{\langle m}^+(0), \dots, e_{\langle m}^+(N-1))^T$ $= (e_m^+(m), \dots, e_m^+(N-1+m))^T$	$\mathbf{e}_{\langle m}^- = (e_{\langle m}^-(0), \dots, e_{\langle m}^-(N-1))^T$ $= (e_m^-(-m), \dots, e_m^-(N-1-m))^T$
	Type (II-f)	Type (II-b)
Least squares estimation	$\alpha'_m = \frac{1}{N} \sum_{n=0}^{N-1} \{e_m^+(n)\}^2$ $\partial \alpha'_m / \partial a_{mi} = 0 \quad (35)$	$\beta'_m = \frac{1}{N} \sum_{n=0}^{N-1} \{x_m^-(n)\}^2$ $\partial \beta'_m / \partial b_{mi} = 0 \quad (37)$
Estimation based on orthogonality principle	$\sum_{n=0}^{N-1} \left\{ \sum_{i=0}^m a_{mi} x(n-i) x(n-j) \right\} = 0$ $(j=1, 2, \dots, m) \quad (36)$	$\sum_{n=0}^{N-1} \left\{ \sum_{i=1}^{m+1} b_{mi} x(n-i) x(n-j) \right\} = 0$ $(j=1, 2, \dots, m) \quad (38)$
Estimation based on Gram-Schmidt procedure	<p style="text-align: center;">—</p> <p style="text-align: right;">(*1)</p>	$\sum_{n=0}^{N-1} x_m^-(n) x_l^-(n) = 0$ $(l \neq m) \quad (39)$
Data vector	<p style="text-align: center;">—</p> <p style="text-align: right;">(*2)</p>	$\mathbf{x}_m = (x(-1-m), \dots, x_m(N-1-m-1))^T$
Prediction error vector	<p style="text-align: center;">—</p> <p style="text-align: right;">(*2)</p>	$\mathbf{x}_m^- = (x_m^-(0), \dots, x_m^-(N-1))^T$

procedure and the other estimation methods is discussed.

2.3 Equivalence Relationship between Three Estimation Procedures

Using the linear prediction theory, it is easy to prove that there is an equivalence relationship between the least squares estimation, estimation based on the orthogonality principle, and estimation based on the Gram-Schmidt procedure.

These three estimation procedures are shown for (I-f), (I-b), (II-f), and (II-b) in Table 1. There are

equivalence relationships between the three estimation procedures in Table 1 (except for type (II-f)). The proofs are omitted, but they are easy to verify.

For estimation based on the Gram-Schmidt procedure for type (II-f) in Table 1, we can specify the following equation:

$$\sum_{n=0}^{N-1} e_m^+(n) e_l^+(n) = 0 \quad (l \neq m). \quad (40)$$

((*1) in Table 1)

Then, the forward prediction error vector $\mathbf{x}_m^+$ can be written as

$$\mathbf{x}_m^+ = (e_m^+(0), e_m^+(1), \dots, e_m^+(N-1))^T. \quad (*2) \text{ in Table 1}$$

It then follows that $g(m)$ in Eq. (20) is constant (zero). However, in the proof⁷⁾ of Theorem 2.1, it was noted that $g(m)$ is not constant. Thus, if $g(m)$ is constant, then it is impossible to find the data vector. Therefore, we cannot obtain $e_m^+(n)$ in Eq. (40) by using the Gram-Schmidt procedure, which means that least squares estimation based on type (II-f) is not correct.

From the results of Gram-Schmidt expansion for type (II-b) in Table 1, the data vector $\mathbf{x}_m$ can be written as

$$\mathbf{x}_m = (x(-1-m), x(-m), \dots, x(N-1-m-1))^T, \quad (41)$$

and the backward prediction error vector $\mathbf{x}_m^-$ can be written as

$$\mathbf{x}_m^- = (x_m^-(0), x_m^-(1), \dots, x_m^-(N-1))^T. \quad (42)$$

It then follows from Eq. (8) that

$$\mathbf{x}_m^- = (e_m^-(-1-m), e_m^-(-m), \dots, e_m^-(N-1-m-1))^T. \quad (43)$$

Comparing Eqs. (23) with (41), and Eqs. (24) with (43), we note that there is a time difference between $\mathbf{x}_{\langle m \rangle}$ and $\mathbf{x}_m$, and between $\mathbf{e}_{\langle m \rangle}^+$ and $\mathbf{x}_m^-$. If we set $h(0) = k(0) = -1$ in Theorem 2.2, then we obtain the same results. It is, however, correct to set $h(0) = k(0) = 0$, that is, $\mathbf{x}_{\langle 0 \rangle} = (x(0), x(1), \dots, x(N-1))^T$, because we want to know the model for the data range $x(0), x(1), \dots, x(N-1)$. In other words, the 0-th backward prediction error range must be identical with the data range $x(0), x(1), \dots, x(N-1)$. Therefore, we should use type (I-b) instead of (II-b).

3. NEW COVARIANCE METHOD (NON-TOEPLITZ FORM METHOD) BASED ON ORTHOGONALIZATION

3.1 Forward Prediction Procedure by Gram-Schmidt Orthogonalization

In the previous section, the dimension of data vectors and prediction error vectors was set to N , and the maximum order of the linear predictor was set to M . The m -th forward prediction error was defined by

$$e_m^+(n) = x(n) - \hat{x}_m(n) \\ = \sum_{i=0}^m a_{mi} x(n-i), \quad a_{m0} = 1$$

for $m=0, 1, \dots, M$. The data vector $\mathbf{x}_{\langle m \rangle}$ was defined by Eq. (17), and the forward prediction error vector $\mathbf{e}_{\langle m \rangle}^+$ was defined by Eq. (18). We now define the forward predictor vector by

$$\mathbf{a}'_m = (a_{mm}, a_{m, m-1}, \dots, a_{m1})^T. \quad (44)$$

Using these vectors, it follows that

$$\mathbf{e}_{\langle m \rangle}^+ = \mathbf{x}_{\langle m \rangle} + \mathbf{X}_{\langle m \rangle} \mathbf{a}'_m, \quad (45)$$

where

$$\mathbf{X}_{\langle m \rangle} = (\mathbf{x}_{\langle 0 \rangle}, \mathbf{x}_{\langle 1 \rangle}, \dots, \mathbf{x}_{\langle m-1 \rangle}). \quad (46)$$

(To be exact, $\mathbf{x}_{\langle m \rangle}$, $\mathbf{e}_{\langle m \rangle}^+$, and $\mathbf{a}'_m$ are represented) (as $\mathbf{x}_{\langle m \rangle, t_0, T}$, $\mathbf{e}_{\langle m \rangle, t_0, T}^+$, and $\mathbf{a}'_{m, t_0, T}$, respectively.)

We now apply the Gram-Schmidt procedure to these data vectors.

(Gram-Schmidt procedure for forward prediction) Initial condition:

$$\mathbf{e}_{\langle 0 \rangle}^+ = \mathbf{x}_{\langle 0 \rangle}. \quad (47-a)$$

For $m=1, 2, \dots, M$,

$$\mathbf{e}_{\langle m \rangle}^+ = \mathbf{x}_{\langle m \rangle} - \sum_{l=0}^{m-1} \gamma_{lm} \mathbf{e}_{\langle l \rangle}^+ \quad (47-b)$$

where

$$\gamma_{lm} = \frac{(\mathbf{e}_{\langle l \rangle}^+, \mathbf{x}_{\langle m \rangle})}{(\mathbf{e}_{\langle l \rangle}^+, \mathbf{e}_{\langle l \rangle}^+)}. \quad (48)$$

Defining

$$\mathbf{E}_{\langle M+1 \rangle} = (\mathbf{e}_{\langle 0 \rangle}^+, \mathbf{e}_{\langle 1 \rangle}^+, \dots, \mathbf{e}_{\langle M \rangle}^+), \quad (49)$$

$$\mathbf{\Gamma}_{M+1} = \begin{bmatrix} 1 & \gamma_{01} & \gamma_{02} & \dots & \gamma_{0M} \\ & 1 & \gamma_{12} & \dots & \gamma_{1M} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \gamma_{M-1, M} \\ 0 & & & & 1 \end{bmatrix}, \quad (50)$$

it follows that

$$\mathbf{X}_{\langle M+1 \rangle} = \mathbf{E}_{\langle M+1 \rangle} \mathbf{\Gamma}_{M+1}. \quad (51)$$

Using the sequence $\{x(n)\}_{n=0}^{N-1+M}$, we obtain coefficients γ_{lm} from Eq. (48), that is,

$$\gamma_{lm} = \frac{\sum_{i=0}^l a_{li} r_{l-i, m}}{\sum_{j=0}^l a_{lj} r_{l-j, l}} \quad \left(\begin{matrix} m=1, 2, \dots, M \\ l=0, 1, \dots, m-1 \end{matrix} \right), \quad (52)$$

where

$$r_{ij} = \frac{1}{N} \sum_{n=0}^{N-1} x(n+i) x(n+j). \quad (53)$$

Assuming that $x(n)$ is zero for $n < 0$ and $n > N-1$, we can define a new forward sample autocorrelation function as follows:

$$r_{ij} = \frac{1}{N} \sum_{n=0}^{N-1-\max(i,j)} x(n+i)x(n+j). \quad (54)$$

In Section 5, we shall discuss the positive definiteness of the autocorrelation matrix whose elements are given by Eq. (54).

Since we have obtained the coefficients $\{\gamma_{lm}\}_{l=0}^{m-1}$, we next determine the forward predictor coefficients.

Theorem 3.1

The forward predictor coefficients are given by

$$\begin{cases} a_{mj} = -\sum_{i=1}^j \gamma_{m-i,m} a_{m-i,j-i} & (m=1, 2, \dots, M) \\ a_{m0} = 1, \quad a_{00} = 1 & (j=1, 2, \dots, m) \end{cases} \quad (55)$$

Proof: It follows from Eq. (47) that

$$e_{\langle m \rangle}^+(n) = x_{\langle m \rangle}(n) - \sum_{l=0}^{m-1} \gamma_{lm} e_{\langle l \rangle}^+(n),$$

that is,

$$e_m^+(n+m) = x(n+m) - \sum_{l=0}^{m-1} \gamma_{lm} e_l^+(n+l). \quad (56)$$

Hence

$$\begin{aligned} \sum_{i=0}^m a_{mi} x(n+m-i) \\ = x(n+m) - \sum_{l=0}^{m-1} \gamma_{lm} \sum_{j=0}^l a_{lj} x(n+l-j). \end{aligned} \quad (57)$$

Comparing both sides of Eq. (57), we obtain Eq. (55). Q.E.D.

Applying Eqs. (52) and (55) alternately to a given autocorrelation function, we obtain the coefficients $\{\gamma_{lm}\}_{l=0}^{m-1}$ and the forward predictor coefficients $\{a_{mi}\}_{i=0}^m$ for $m=1, 2, \dots, M$.

3.2 Backward Prediction Procedure by Gram-Schmidt Orthogonalization

In the same manner as for forward prediction, we can obtain a backward prediction procedure. The data vector $\mathbf{x}_{\langle m \rangle}$ and the backward prediction error vector $\mathbf{e}_{\langle m \rangle}^-$ were defined by Eqs. (23) and (24), respectively. We now define the backward predictor vector by

$$\mathbf{b}_m = (b_{m1}, b_{m2}, \dots, b_{mm})^T. \quad (58)$$

Using these vectors, we have

$$\mathbf{e}_{\langle m \rangle}^- = \mathbf{x}_{\langle m \rangle} + \mathbf{X}_{\langle m \rangle} \mathbf{b}_m, \quad (59)$$

where

$$\mathbf{X}_{\langle m \rangle} = (\mathbf{x}_{\langle 0 \rangle}, \mathbf{x}_{\langle 1 \rangle}, \dots, \mathbf{x}_{\langle m-1 \rangle}). \quad (60)$$

(Gram-Schmidt procedure for backward prediction)

Initial condition:

$$\mathbf{e}_{\langle 0 \rangle}^- = \mathbf{x}_{\langle 0 \rangle}. \quad (61-a)$$

For $m=1, 2, \dots, M$,

$$\mathbf{e}_{\langle m \rangle}^- = \mathbf{x}_{\langle m \rangle} - \sum_{l=0}^{m-1} \delta_{lm} \mathbf{e}_{\langle l \rangle}^- \quad (61-b)$$

where

$$\delta_{lm} = \frac{(\mathbf{e}_{\langle l \rangle}^-, \mathbf{x}_{\langle m \rangle})}{(\mathbf{e}_{\langle l \rangle}^-, \mathbf{e}_{\langle l \rangle}^-)}. \quad (62)$$

Defining

$$\mathbf{E}_{\langle M+1 \rangle} = (\mathbf{e}_{\langle 0 \rangle}^-, \mathbf{e}_{\langle 1 \rangle}^-, \dots, \mathbf{e}_{\langle M \rangle}^-), \quad (63)$$

we have

$$\mathbf{X}_{\langle M+1 \rangle} = \mathbf{E}_{\langle M+1 \rangle} \mathbf{A}_{M+1}, \quad (64)$$

where $\mathbf{A}_{M+1}$ is an upper triangular matrix with all diagonal elements equal to unity.

Using the sequence $\{x(n)\}_{n=-M}^{N-1}$, we can obtain the coefficients δ_{lm} in the same manner as γ_{lm} .

$$\delta_{lm} = \frac{\sum_{i=1}^{l+1} b_{li} r_{1-i, -m}}{\sum_{j=1}^{l+1} b_{lj} r_{1-j, -l}}, \quad \begin{cases} m=1, 2, \dots, M \\ l=0, 1, \dots, m-1 \end{cases} \quad (65)$$

where

$$r_{-i, -j} = \frac{1}{N} \sum_{n=0}^{N-1} x(n-i)x(n-j). \quad (66)$$

Assuming that $x(n)$ is zero for $n < 0$ and $n > N-1$, we can define a new backward sample autocorrelation function as follows:

$$r_{-i, -j} = \frac{1}{N} \sum_{n=\max(i,j)}^{N-1} x(n-i)x(n-j). \quad (67)$$

Since we have obtained the coefficients $\{\delta_{lm}\}_{l=0}^{m-1}$, we next determine the backward predictor coefficients by Theorem 3.2. The proof of Theorem 3.2 is similar to that for Theorem 3.1.

Theorem 3.2

The backward predictor coefficients are given by

$$\begin{cases} b_{mj} = -\sum_{i=j-1}^{m-1} \delta_{im} b_{ij} & (m=1, 2, \dots, M) \\ b_{m, m+1} = 1 & (j=1, 2, \dots, m) \end{cases} \quad (68)$$

Applying Eqs. (65) and (68) alternately to a given autocorrelation function, we obtain the coefficients

$\{\delta_{lm}\}_{l=0}^{m-1}$ and the backward predictor coefficients $\{b_{mi}\}_{i=1}^{m+1}$, for $m=1, 2, \dots, M$.

The forward and backward prediction errors are shown in Ref. 7). Their digital filter representations are shown in Figs. 2 and 3 in Ref. 7).

4. NEW COVARIANCE METHOD (NON-TOEPLITZ FORM METHOD) USING NORMAL EQUATIONS

The previous section showed prediction procedures using Gram-Schmidt orthogonalization. This section shows prediction procedures using normal equations, by using matrix decomposition.

4.1 Forward Prediction Procedure Using Normal Equation

From Eq. (45), the linear model is represented as

$$\mathbf{x}_{(m)} = -\mathbf{X}_{(m)} \mathbf{a}'_m + \mathbf{e}_{(m)}^+ . \quad (69)$$

Then the least squares predictor coefficients can be obtained by solving

$$(\mathbf{X}_{(m)}^T \mathbf{X}_{(m)}) \mathbf{a}'_m = -\mathbf{X}_{(m)}^T \mathbf{x}_{(m)} . \quad (70)$$

Defining

$$\mathbf{R}_m^+ = \mathbf{X}_{(m)}^T \mathbf{X}_{(m)} / N , \quad \mathbf{r}_m^+ = \mathbf{X}_{(m)}^T \mathbf{x}_{(m)} / N , \quad (71)$$

we have

$$\mathbf{R}_m^+ \mathbf{a}'_m = -\mathbf{r}_m^+ . \quad (72)$$

Note that this equation is equivalent to Eq. (10).

For $m=M$,

$$\mathbf{R}_M^+ \mathbf{a}'_M = -\mathbf{r}_M^+ , \quad (73)$$

where

$$\mathbf{R}_M^+ = [r_{ij}^+]_{i,j=0}^{M-1} , \quad (74-a)$$

$$\mathbf{r}_M^+ = [r_{0M}, r_{1M}, \dots, r_{M-1,M}]^T , \quad (74-b)$$

$$\mathbf{a}'_M = [a_{MM}, a_{M,M-1}, \dots, a_{M1}]^T . \quad (74-c)$$

The predictor coefficients $\{a_{mi}\}_{i=1}^M$ can be obtained by solving Eq. (73).

It follows from Eq. (51) that

$$\begin{aligned} \mathbf{R}_M^+ &= \mathbf{X}_{(M)}^T \mathbf{X}_{(M)} / N \\ &= \mathbf{\Gamma}_M^T \mathbf{E}_{(M)}^T \mathbf{E}_{(M)} \mathbf{\Gamma}_M / N \\ &= \mathbf{\Gamma}_M^T \mathbf{D}_M^+ \mathbf{\Gamma}_M / N \\ &= \mathbf{W}_M \mathbf{\Gamma}_M , \end{aligned}$$

where $\mathbf{D}_M^+$ is a diagonal matrix. Therefore, note that by decomposing $\mathbf{R}_M^+$ by LU transformation (L is a lower triangular matrix, U an upper triangular matrix with all diagonal elements equal to unity),⁸⁾

the coefficients $\{\gamma_{lm}\}_{l=0}^{m-1}$ in Eq. (52) can be obtained.

We can rewrite Eq. (73) as

$$\mathbf{R}_M^+ \mathbf{a}'_M = -\mathbf{r}_M^+ , \quad (75)$$

where

$$\mathbf{R}_M^+ = [r_{ij}^+]_{i,j=1}^M , \quad r_{ij}^+ = r_{M-j,M-i} , \quad (76-a)$$

$$\mathbf{r}_M^+ = [r_{M,M-1}, r_{M,M-2}, \dots, r_{M0}]^T , \quad (76-b)$$

$$\mathbf{a}'_M = [a_{M1}, a_{M2}, \dots, a_{MM}]^T . \quad (76-c)$$

We now define the m -dimensional equation related to Eq. (75) as follows:

$$\mathbf{R}_m^{(M)} \mathbf{a}_{mm}^{(M)} = -\mathbf{r}_m^{(M)} , \quad * \quad (77)$$

where $\mathbf{R}_m^{(M)}$ is a principal minor of $\mathbf{R}_M^+$,

$$\mathbf{R}_m^{(M)} = [r_{ij}^+]_{i,j=1}^m , \quad r_{ij}^+ = r_{M-j,M-i} , \quad (78)$$

$$\mathbf{r}_m^{(M)} = [r_1^+, r_2^+, \dots, r_m^+]^T , \quad (79)$$

$$r_i^+ = r_{M,M-i} ,$$

$$\mathbf{a}_{mm}^{(M)} = [a_{m1}^{(M)}, a_{m2}^{(M)}, \dots, a_{mm}^{(M)}]^T . \quad (80)$$

Then, it follows by LU transformation that

$$\mathbf{R}_m^{(M)} = \mathbf{W}_m' \mathbf{\Gamma}_m' , \quad (81)$$

where

$$\mathbf{W}_m' = \begin{bmatrix} w'_{11} & & & \\ w'_{21} & w'_{22} & & \mathbf{0} \\ \vdots & & & \\ w'_{m1} & w'_{m2} & \dots & w'_{mm} \end{bmatrix} ,$$

$$\mathbf{\Gamma}_m' = \begin{bmatrix} 1 & \gamma'_{01} & \dots & \gamma'_{0,m-1} \\ & 1 & \dots & \gamma'_{1,m-1} \\ & & \ddots & \vdots \\ & \mathbf{0} & & 1 \end{bmatrix} .$$

Defining

$$\boldsymbol{\kappa}_{mm}^{(M)} = [\kappa_{m1}^{(M)}, \kappa_{m2}^{(M)}, \dots, \kappa_{mm}^{(M)}]^T , \quad (82)$$

we obtain from Eqs. (77) and (81)

$$\left\{ \begin{array}{l} \mathbf{W}_m' \boldsymbol{\kappa}_{mm}^{(M)} = -\mathbf{r}_m^{(M)} \end{array} \right. \quad (83-a)$$

$$\left\{ \begin{array}{l} \mathbf{\Gamma}_m' \mathbf{a}_{mm}^{(M)} = \boldsymbol{\kappa}_{mm}^{(M)} . \end{array} \right. \quad (83-b)$$

We now solve Eq. (83) to obtain the set of solutions $\{\kappa_{mi}^{(M)}\}$, to determine the relation between the auxiliary variable κ_{Mm} and the PARCOR coefficient k_m . The following Lemma 4.1 can be obtained by LU transformation.⁹⁾

* It follows from Eqs. (75) to (77) that

$$\mathbf{a}_{MM}^{(M)} = \mathbf{a}_M , \quad \mathbf{R}_M^{(M)} = \mathbf{R}_M^+ , \quad \mathbf{r}_M^{(M)} = \mathbf{r}_M^+ .$$

Lemma 4.1

Let $\mathbf{R}_M^+$ be a nonsingular matrix, and $\mathbf{R}_m^{(M)}$ a nonsingular matrix for $m=1, 2, \dots, M-1$. Supposing that

$$\mathbf{R}_{m-1}^{(M)} = \mathbf{W}_{m-1}' \mathbf{I}_{m-1}',$$

is obtained by LU transformation and $\mathbf{K}_{m-1,m-1}^{(M)}$ is known, then

$$\mathbf{K}_{m,m-1}^{(M)} = \mathbf{K}_{m-1,m-1}^{(M)} \quad (84)$$

and $\mathbf{K}_{m,m}^{(M)}$ can be obtained recursively. Setting

$$\begin{aligned} \mathbf{K}_{Mi} &= (\mathbf{K}_{M1}, \mathbf{K}_{M2}, \dots, \mathbf{K}_{Mi}) \\ &= \mathbf{K}_{mi}^{(M)} \quad (1 \leq i \leq m), \quad m=1, 2, \dots, M, \end{aligned}$$

we have

$$\begin{aligned} \mathbf{K}_{Mm} &= \mathbf{K}_{mm}^{(M)} \\ &= \frac{-\mathbf{r}_{M,M-m} - \sum_{i=1}^{m-1} \mathbf{W}_{mi}' \mathbf{K}_{Mi}}{\mathbf{W}_{mm}'} \quad (85) \end{aligned}$$

Proof: The proof is shown in Ref. 7). Q.E.D.

We shall obtain solution Eq. (83). It follows from Eq. (83-b) that

$$\begin{bmatrix} \mathbf{I}_{m-1}' & \mathbf{r}_{m-1}' \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_{m,m-1}^{(M)} \\ \mathbf{a}_{m,m}^{(M)} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{M,m-1} \\ \mathbf{K}_{Mm} \end{bmatrix}, \quad (86)$$

where $\mathbf{r}_{m-1}'$ is given by LU transformation.

Theorem 4.1

If $\mathbf{K}_{Mm}$ and $\mathbf{r}_{m-1}'$ are given, then we can obtain $\mathbf{a}_{m,m}^{(M)}$ recursively, as follows:

$$\begin{cases} \mathbf{a}_{mm}^{(M)} = \mathbf{K}_{Mm} \\ \mathbf{a}_{mi}^{(M)} = \mathbf{K}_{Mi} - \sum_{j=i+1}^m \gamma_{i-1,j-1}' \mathbf{a}_{mj}^{(M)}. \end{cases} \quad (87-a)$$

$$(87-b)$$

Proof: The proof is shown in Ref. 7). Q.E.D.

(In general, $\mathbf{a}_{mi}^{(M)} \triangleq \mathbf{a}_{mi}$ ($m \neq M$) and $\mathbf{a}_{Mi}^{(M)} = \mathbf{a}_{Mi}$.)

The solution $\mathbf{a}_M$ of Eq. (75) is given by setting $m=M$ in Eq. (87).

Using Eq. (75) instead of Eq. (73) as the normal equation, it follows from Eq. (87-a) that if input $X(t)$ is a weakly stationary process (the autocorrelation matrix is Toeplitz form), then the auxiliary variable $\mathbf{K}_{Mm}$ is equal to PARCOR coefficient k_m for $m=1, 2, \dots, M$. However, it should be noted that if the input is not stationary, then this variable $\mathbf{K}_{Mm}$ is not equal, in general, to the coefficient k_m .

4.2 The Backward Prediction Procedure Using Normal Equation

In the same manner as for forward prediction, we

can obtain the backward prediction procedure. The linear model and the minimum mean square estimation are represented by

$$\mathbf{x}_{\langle m \rangle} = -\mathbf{X}_{\langle m \rangle} \mathbf{b}_m + \mathbf{e}_{\langle m \rangle}^-, \quad (88)$$

$$(\mathbf{X}_{\langle m \rangle}^T \mathbf{X}_{\langle m \rangle}) \mathbf{b}_m = -\mathbf{X}_{\langle m \rangle}^T \mathbf{x}_{\langle m \rangle}^-, \quad (89)$$

respectively. Defining

$$\mathbf{R}_m^- = \mathbf{X}_{\langle m \rangle}^T \mathbf{X}_{\langle m \rangle} / N, \quad \mathbf{r}_m^- = \mathbf{X}_{\langle m \rangle}^T \mathbf{x}_{\langle m \rangle}^- / N, \quad (90)$$

we have

$$\mathbf{R}_m^- \mathbf{b}_m = -\mathbf{r}_m^-.$$

Note that this equation is equivalent to Eq. (12). For $m=M$,

$$\mathbf{R}_M^- \mathbf{b}_M = -\mathbf{r}_M^-, \quad (91)$$

where

$$\mathbf{R}_M^- = [\mathbf{r}_{-i, -j}^-]_{i,j=0}^{M-1}, \quad (92-a)$$

$$\mathbf{r}_M^- = [\mathbf{r}_{0, -M}^-, \mathbf{r}_{-1, -M}^-, \dots, \mathbf{r}_{-M+1, -M}^-]^T, \quad (92-b)$$

$$\mathbf{b}_M = [\mathbf{b}_{M1}, \mathbf{b}_{M2}, \dots, \mathbf{b}_{MM}]^T. \quad (92-c)$$

The predictor coefficients $\{\mathbf{b}_{Mi}\}_{i=1}^M$ can be obtained by solving Eq. (91). It follows from Eqs. (64) and (90) that

$$\begin{aligned} \mathbf{R}_M^- &= \mathbf{A}_M^T \mathbf{E}_{\langle M \rangle}^T \mathbf{E}_{\langle M \rangle} \mathbf{A}_M / N \\ &= \mathbf{A}_M^T \mathbf{D}_M^- \mathbf{A}_M / N \\ &= \Phi_M \mathbf{A}_M, \end{aligned} \quad (93)$$

where $\mathbf{D}_M^-$ is a diagonal matrix. Decomposing $\mathbf{R}_M^-$ by LU transformation, one can obtain the coefficients $\{\delta_{im}\}_{i=0}^{m-1}$ in Eq. (65).

We can rewrite Eq. (91) as

$$\mathbf{R}_M^- \mathbf{b}_M' = -\mathbf{r}_M'^-, \quad (94)$$

where

$$\mathbf{R}_M^- = [\mathbf{r}_{ij}']_{i,j=1}^M = [\mathbf{r}_{-M+j, -M+i}']_{i,j=1}^M, \quad (95-a)$$

$$\mathbf{r}_M'^- = [\mathbf{r}_{-M, -M+1}', \mathbf{r}_{-M, -M+2}', \dots, \mathbf{r}_{-M, 0}']^T, \quad (95-b)$$

$$\mathbf{b}_M' = [\mathbf{b}_{MM}, \mathbf{b}_{M,M-1}, \dots, \mathbf{b}_{M1}]^T. \quad (95-c)$$

In the same manner as for forward prediction, we can say that if input $X(t)$ is a weakly stationary process, then the auxiliary variables in the matrix decomposition are equivalent to PARCOR coefficients.⁷⁾

5. POSITIVE DEFINITENESS OF THE AUTOCORRELATION MATRIX

In this section the positive definiteness of the autocorrelation matrices is discussed. It is shown that

the autocorrelation matrices whose elements are given by the autocorrelation functions proposed by others are not positive definite. However, the matrices whose elements are given by the autocorrelation functions proposed in this paper are positive definite.

5.1 Positive Definiteness of the Forward Autocorrelation Matrix

In Section 3, the new sample autocorrelation function was defined as Eq. (54). In Theorem 5.1, it is shown that the matrix composed of Eq. (54) is positive definite. In Theorem 5.2, it is shown that if the matrix is positive definite, then the forward mean square error α_m is positive, and the converse is also true. Then, we can determine the solution $\{a_{mi}\}_{i=1}^m$ uniquely. Theorem 5.1

Let $\mathbf{R}_{M+1}^+$ be a $(M+1) \times (M+1)$ autocorrelation matrix whose elements are given by the autocorrelation function defined by

$$r_{ij} = \frac{1}{N} \sum_{n=0}^{N-1-\max(i,j)} x(n+i)x(n+j). \quad (54)$$

Then $\mathbf{R}_{M+1}^+$ is positive definite.

Proof: Consider a quadratic form

$$\mathbf{c}_{M+1} \mathbf{R}_{M+1}^+ \mathbf{c}_{M+1}^T, \quad (96)$$

where

$$\mathbf{c}_{M+1} = [c_0, c_1, \dots, c_M]$$

is a $1 \times (M+1)$ vector and is not the zero vector, and $\mathbf{R}_{M+1}^+$ is derived by Eq. (74-a). Then

$$\begin{aligned} \mathbf{c}_{M+1} \mathbf{R}_{M+1}^+ \mathbf{c}_{M+1}^T &= c_0^2 r_{00} + c_1^2 r_{11} + \dots + c_M^2 r_{MM} \\ &\quad + 2c_0 c_1 r_{01} + 2c_0 c_2 r_{02} + \dots + 2c_0 c_M r_{0M} \\ &\quad \vdots \\ &\quad + 2c_{M-2} c_{M-1} r_{M-2, M-1} + 2c_{M-2} c_M r_{M-2, M} \\ &\quad + 2c_{M-1} c_M r_{M-1, M}. \end{aligned} \quad (97)$$

Substituting Eq. (54) into Eq. (97), we obtain

$$\begin{aligned} \mathbf{c}_{M+1} \mathbf{R}_{M+1}^+ \mathbf{c}_{M+1}^T &= \sum_{n=0}^{N-M-1} (c_0 x_n + c_1 x_{n+1} + \dots + c_M x_{n+M})^2 \\ &\quad + \{(c_0 x_{N-M} + c_1 x_{N-M+1} + \dots + c_M x_{N-1})^2 \\ &\quad + (c_0 x_{N-M+1} + c_1 x_{N-M+2} + \dots + c_{M-2} x_{N-1})^2 \\ &\quad \vdots \\ &\quad + (c_0 x_{N-2} + c_1 x_{N-1})^2 \\ &\quad + (c_0 x_{N-1})^2\}. \end{aligned} \quad (98)$$

Equation (98) is nonnegative for nontrivial $c_0, c_1, \dots, c_M$. Supposing $\mathbf{c}_{M+1} \mathbf{R}_{M+1}^+ \mathbf{c}_{M+1}^T \cdot N = 0$, we obtain

$$x_{N-1} = x_{N-2} = \dots = x_{N-M} = 0, \quad (\text{due to the second term of Eq. (98)})$$

and

$$x_{N-M-1} = x_{N-M-2} = \dots = x_0 = 0. \quad (\text{due to the first term of Eq. (98)})$$

Therefore, Eq. (98) must be positive for nontrivial $c_0, c_1, \dots, c_M$. Thus, $\mathbf{R}_{M+1}^+$ is positive definite.

Q.E.D.

$$\left(\begin{array}{l} \text{Substituting Eq. (53) to Eq. (97), we obtain} \\ \mathbf{c}_{M+1} \mathbf{R}_{M+1}^+ \mathbf{c}_{M+1}^T \cdot N \\ = \sum_{n=0}^{N-1} (c_0 x_n + c_1 x_{n+1} + \dots + c_M x_{n+M})^2. \\ \text{It is clear that this equation is nonnegative.} \end{array} \right)$$

Theorem 5.2

Let $\mathbf{R}_{M+1}^+$ be an autocorrelation matrix, and α_m a forward sample mean square error defined by

$$\alpha_m = \frac{1}{N} \sum_{n=0}^{N-1} \{e_{(m)}^+(n)\}^2. \quad (29)$$

If $\mathbf{R}_{M+1}^+$ is positive definite, then

$$\alpha_m > 0$$

for $m=0, 1, \dots, M$. The converse is also true.

Proof: The proof is shown in Ref. 7).

It follows from Eqs. (48) and (52) that

$$\gamma_{lm} = \frac{\sum_{i=0}^l a_{li} r_{l-i, m}}{\alpha_l} \quad \left(\begin{array}{l} m=1, 2, \dots, M \\ l=0, 1, \dots, m-1 \end{array} \right). \quad (99)$$

It should be noted that if $\alpha_l = 0$ for some l , then one cannot obtain the predictor coefficients $\{a_{mi}\}_{i=1}^m$ for $m \geq l$. Therefore, it is important that $\mathbf{R}_{M+1}^+$ be positive definite.

5.2 Positive definiteness of the Backward Autocorrelation Matrix

In Theorem 5.3, it is shown that the matrix whose elements are given by autocorrelation function Eq. (67) is positive definite. In Theorem 5.4, it is shown that if the matrix is positive definite, then the backward mean square error β_m is positive, and the converse is also true. Then, we can determine the solution $\{b_{mi}\}_{i=1}^m$ uniquely. Theorems 5.3 and 5.4 are proved in the same manner as Theorems 5.1 and 5.2, respectively.

Theorem 5.3

Let $\mathbf{R}_{M+1}^-$ be an $(M+1) \times (M+1)$ autocorrelation matrix whose elements are given by the autocorrelation function defined by

$$r_{-i,-j} = \frac{1}{N} \sum_{n=\max(i,j)}^{N-1} x(n-i)x(n-j). \quad (67)$$

Then $\mathbf{R}_{M+1}^-$ is positive definite.

Theorem 5.4

Let $\mathbf{R}_{M+1}^-$ be an autocorrelation matrix, and β_m a forward sample mean square error defined by

$$\beta_m = \frac{1}{N} \sum_{n=0}^{N-1} \{e_{\langle m \rangle}^-(n)\}^2. \quad (32)$$

If $\mathbf{R}_{M+1}^-$ is positive definite, then

$$\beta_m > 0$$

for $m=0, 1, \dots, M$. The converse is also true.

It follows from Eqs. (62) and (65) that

$$\delta_{lm} = \frac{\sum_{i=1}^{l+1} b_{li} r_{1-i,-m}}{\beta_l} \quad \left(\begin{array}{l} m=1, 2, \dots, M \\ l=0, 1, \dots, m-1 \end{array} \right). \quad (100)$$

To determine the backward predictor coefficients, it is important that $\beta_m > 0$ ($m=0, 1, 2, \dots, M$), that is, $\mathbf{R}_{M+1}^-$ is positive definite.

6. DISCUSSION

6.1 Discussion on the Covariance Method Based on Orthogonalization

The non-Toeplitz form method proposed in this paper and the covariance methods proposed by others are compared.

(i) In the non-Toeplitz form method, there is no time difference between the data and its predictors, because the backward prediction model was defined by Eqs. (4), (5) and (6).

(ii) To obtain error vectors from data vectors by using the Gram-Schmidt orthogonalization procedure, the forward and backward data vectors were defined as Eqs. (17) and (23); the forward and backward prediction error vectors were defined as Eqs. (18) and (24).

(iii) The autocorrelation matrix whose elements are given by

$$r_{ij} = \frac{1}{N} \sum_{n=0}^{N-1} x(n+i)x(n+j) \quad (53)$$

is nonnegative definite. Then, it was shown in Section 5.1 that the solution $\{a_{mi}\}_{i=1}^m$ cannot always be obtained uniquely. On the other hand, the autocorrelation

matrix whose elements are given by

$$r_{ij} = \frac{1}{N} \sum_{n=0}^{N-1-\max(i,j)} x(n+i)x(n+j) \quad (54)$$

proposed in this paper is positive definite. Then, the solution $\{a_{mi}\}_{i=1}^m$ can always be obtained uniquely.

(iv) In the same manner as for forward prediction, it was shown that the autocorrelation matrix whose elements are given by

$$r_{-i,-j} = \frac{1}{N} \sum_{n=\max(i,j)}^{N-1} x(n-i)x(n-j) \quad (67)$$

proposed in this paper is positive definite. Then, the solution $\{b_{mi}\}_{i=1}^m$ can always be obtained uniquely.

(v) The autocorrelation matrix whose elements are given by the backward sample autocorrelation function used by Markel and Gray (Ref. 4), p. 51 *et al.*; that is

$$r_{-i,-j}^* = \frac{1}{N} \sum_{n=M}^{N-1} x(n-i)x(n-j) \quad (101)$$

is nonnegative definite.

(vi) Markel and Gray (Ref. 4), p. 55) also derived the coefficient

$$\delta_{lm}^* = \frac{\sum_{i=1}^{l+1} b_{li} r_{1-i,-m}^*}{\sum_{j=1}^{l+1} b_{lj} r_{-j,-l-1}^*} \quad \left(\begin{array}{l} m=1, 2, \dots, M \\ l=0, 1, \dots, m-1 \end{array} \right) \quad (102)$$

in a backward prediction procedure. There is a time difference between Eqs. (102) and (65), because there is a time difference between the prediction error $e_m^-(n)$ defined in this paper and $x_m^-(n)$ used by Markel and Gray (shown in Section 2.1). According to Theorem 2.2 and Section 2.3, Eq. (65) derived in this paper should be used.

(vii) In addition, Markel and Gray use Levinson's recurrence formula;

$$\begin{cases} a_{m+1,i} = a_{mi} + k_{m+1} b_{mi} \\ k_{m+1} = a_{m+1,m+1} \end{cases}$$

that can be used when the input $X(t)$ is weakly stationary (its sample autocorrelation function is a one variable function and depends only on time difference). This contradicts the assumption that the autocorrelation function in the covariance method is a two variable function.

6.2 Discussion on the Covariance Method Using Normal Equations

(i) For forward prediction, we used Eq. (75) in-

stead of Eq. (73) as the normal equation. It follows from Eq. (87-a) that if input $X(t)$ is weakly stationary, then the auxiliary variables $\{\kappa_{Mm}\}$ in Section 4 are equivalent to PARCOR coefficients $\{k_m\}$. But the coefficients $\{\gamma_{lm}\}_{l=0}^{m-1}$ cannot be obtained by decomposing R_M^+ in Eq. (75). If we want to obtain the coefficients, we must decompose R_M^+ in Eq. (73). A similar result was obtained for backward prediction.

(ii) In recent papers, some methods for solving normal equations by using matrix decomposition were proposed. As an autocorrelation function, Morf *et al.*³⁾ used

$$r_{ij} = \sum_{n=M}^{N-1} x(n-i)x(n-j). \quad (103)$$

Friedlander⁶⁾ used an algorithm which is equivalent to

$$r_{ij} = \sum_{n=0}^{N-1-M} x(n+i)x(n+j), \quad (104)$$

(for the covariance method in Ref. 6)). However, the autocorrelation matrices whose elements are given by Eqs. (103) and (104) are nonnegative definite. Hence the solution $\{a_{Mi}\}_{i=1}^M$ cannot always be obtained uniquely.

(iii) The main difference between the covariance method (non-Toeplitz form method) based on Gram-Schmidt orthogonalization and the covariance method (non-Toeplitz form method) using normal equations is as follows. If we use the covariance method using normal equations, we obtain only $\{a_{Mi}\}_{i=1}^M$ or $\{b_{Mi}\}_{i=1}^M$ from an M -th order normal equation. However, if we use the covariance method based on Gram-Schmidt orthogonalization, we can obtain $\{a_{mi}\}_{i=1}^m$ or $\{b_{mi}\}_{i=1}^m$ for all m , that is, $m=1, 2, \dots, M$.

7. CONCLUDING REMARKS

The following results have been obtained.

In Section 2, the definitions of forward and backward prediction were given. It was shown that the definitions presented in this paper are unique.

In Section 3, the new forward and backward prediction procedures based on orthogonalization were presented. The new forward and backward sample autocorrelation functions were defined.

In Section 4, the new covariance methods using normal equations were presented. Given that the input $X(t)$ is stationary, it was shown that PARCOR coefficients $\{k_m\}_{m=1}^M$ are equal to the auxiliary vari-

ables $\{\kappa_{Mm}\}_{m=1}^M$.

In Section 5, it was proved that the autocorrelation matrices whose elements are given by the autocorrelation functions proposed in this paper are positive definite. The relationship between the positive definiteness of the autocorrelation matrix and the positiveness of the sample mean square error was also shown.

In Section 6, the new covariance method (non-Toeplitz form method) proposed in this paper and the covariance methods proposed by others were compared.

We now have an accurate analysis method, the non-Toeplitz form method proposed in this paper, for speech analysis-synthesis without the weak points discussed in Section 1.

ACKNOWLEDGMENTS

The author would like to acknowledge the guidance of Dr. Tosio Kitagawa, the President of IIAS-SIS. The author also wishes to express his hearty thanks to Dr. Akio Tanaka for helpful discussion, and Dr. Koji Tajima and Dr. Mitsuhiko Toda for their helpful comments.

REFERENCES

- 1) F. Itakura, "Extraction of feature parameters of speech by statistical method," Proc. 8th Symp. Speech Inf. Process., RICE Tohoku Univ., Sendai, Japan, II.5.1-II.5.12 (1971) (in Japanese).
- 2) B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of speech wave," J. Acoust. Soc. Am. **50**, 637-655 (1971).
- 3) M. Morf, B. Dickinson, T. Kailath, and A. Vieira, "Efficient solution of covariance equations for linear prediction," IEEE Trans. ASSP-**25**, 429-433 (1977).
- 4) J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech* (Springer-Verlag, Berlin, 1976), Chap. 2-3.
- 5) S. Chandra and W. Lin, "Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis," IEEE Trans. ASSP-**22**, 403-415 (1974).
- 6) B. Friedlander, "Lattice methods for spectral estimation," IEEE Proc. **70**, 990-1017 (1982).
- 7) M. Komura, "Covariance methods for linear prediction of speech," Res. Rep. IIAS-SIS Fujitsu Ltd., Jpn. **41**, 1-40 (1983).
- 8) S. Makinouchi and T. Torii, *Numerical Analysis*, (Ohm, Tokyo, 1973), Chap. 2 (in Japanese).
- 9) M. Komura and A. Tanaka, "Covariance methods in linear prediction of speech," IECEJ Tech. Rep. CAS82-182, 79-86 (1983) (in Japanese).