Ann. Inst. Statist. Math.
Vol. 44, No. 2, 239–260 (1992)

IDENTIFICATION OF NON-MINIMUM PHASE TRANSFER FUNCTION USING HIGHER-ORDER SPECTRUM

MASAYUKI KUMON

Department of Artificial Intelligence, Kyushu Institute of Technology, 680-4 Ohaza-kawazu, Iizuka, Fukuoka 820, Japan

(Received March 23, 1990; revised June 15, 1991)

Abstract. The present paper treats the identification of parametric non-minimum phase transfer function. We propose a method of identification based on the inner outer factorization of stable transfer function. It consists of identifying the outer and inner parts of a transfer function separately. The outer part is identified by the use of the second-order spectral estimate from the observed linear process, while the inner part is identified by the use of a higher-order cumulant spectral estimate from the observed process. Respective parameter estimators are determined in the light of asymptotic efficiency. In order to estimate the order of the inner part of a transfer function, a criterion is proposed. It is introduced based on the same principle as in the case of Akaike's AIC.

Key words and phrases: All-pass, asymptotic efficiency, cumulant spectrum, inner function, linear process, minimum phase, non-Gaussian, outer function.

1. Introduction

The present paper deals with the stationary time series which is generated as a linear process. Assume that the random variables $\{e_t\}$, $t = 0, \pm 1, \pm 2, \dots$, are independent and identically distributed with mean zero $E\, e_t = 0$. Let $\{h_j\}$, $j = 0, 1, 2, \dots$, be a sequence of real constants with $\sum_{j=0}^{\infty} h_j^2 < \infty$. The linear process $\{X_t\}$ generated by $\{h_j\}$ and $\{e_t\}$ is given by

$$X_t = \sum_{j=0}^{\infty} h_j e_{t-j}. \tag{1.1}$$

This scheme is regarded as a linear system with output X_t being driven by e_t through a linear filter with impulse response $\{h_j\}$. Let $H(z) = \sum_{j=0}^{\infty} h_j z^j$ be the z-transform corresponding to the process $\{X_t\}$. Then, the frequency response function or transfer function of the linear filter is

$$H(\lambda) = \sum_{j=0}^{\infty} h_j \exp(-ij\lambda). \tag{1.2}$$

We are concerned with the estimation of $H(\lambda)$ on the basis of observations only on the process $\{X_t\}$. This is a standard problem called system identification, and many other results have already been accumulated until 1980 (see e.g. Hannan (1973), Box and Jenkins (1976), Kabaila (1980)). However, these results are mostly based on assumption that the distribution of $\{e_t\}$ is Gaussian. If normality is not assumed, these results are based on the second-order spectrum of $\{X_t\}$, and the two approaches are essentially same.

Under the Gaussian assumption, the full probability structure of $\{X_t\}$ is determined by its second-order spectral density given by

$$f(\lambda) = (2\pi)^{-1} K_2 |H(\lambda)|^2, \tag{1.3}$$

where K_2 denotes the second-order cumulant or the variance of $\{e_t\}$. Therefore, in the identification of $H(\lambda)$ based on $\{X_t\}$, any information about the phase of $H(\lambda)$ can not be obtained in the Gaussian case. If $H(z)$ is a rational function $H(z) = g(z)/f(z)$ with $f(z)$, $g(z)$ polynomials

$$f(z) = \sum_{k=0}^{p} a_k z^k, \quad a_0 = 1, \qquad g(z) = \sum_{k=0}^{q} b_k z^k, \quad b_0 \neq 0,$$

the process $\{X_t\}$ is a finite parameter autoregressive moving average (ARMA) process,

$$\sum_{j=0}^{p} a_j X_{t-j} = \sum_{k=0}^{q} b_k e_{t-k}.$$

In the Gaussian case, it is the custom to assume that all the roots of $f(z)$ and $g(z)$ are outside the unit disk $|z| \leq 1$ in the complex plane. As for $f(z)$, this assumption has a physical meaning, since it implies the stability of the system. However, the assumption on the roots of $g(z)$, which is often called the invertibility condition, has nothing to do with the true structure of the system.

We sometimes regard the z-transform $H(z) = \sum_{j=0}^{\infty} h_j z^j$ as a function of the complex variable z. A transfer function $H(z)$ is said to be stable, when $H(z)$ is analytic in the unit disk $|z| < 1$. Then, the class of stable transfer functions $H(z) = \sum_{j=0}^{\infty} h_j z^j$ with $\sum_{j=0}^{\infty} h_j^2 < \infty$ corresponds to the class H^2 in the theory of H^p spaces (see Duren ((1970), Section 1.1)). It is known that every function $H(z)$ of class H^p $(p > 0)$ has a unique factorization of the form

$$H(z) = A(z)B(z), \tag{1.4}$$

where $A(z)$ is called an outer function, which is stable, i.e. analytic in $|z| < 1$, and has no zeros in $|z| < 1$. $B(z)$ is called an inner function, which is stable, having the properties $|B(z)| < 1$ in $|z| < 1$ and $|B(z)| = 1$ on $|z| = 1$. This is called the inner outer factorization or the canonical factorization (see Duren ((1970), Section 2.4)). Linear filters with transfer functions $A(\lambda)$ and $B(\lambda)$ are called the minimum phase and the all-pass filters.

These observations show that the identification of the inner part is impossible under the Gaussian assumption. For the purpose of full identification of the transfer function, we assume that the distribution of $\{e_t\}$ is non-Gaussian. But we do not need to know the actual non-Gaussian distribution of $\{e_t\}$. In the present paper, the identification is made by using the second- and a higher-order spectral estimates from $\{X_t\}$ which do not require this knowledge.

Identifications using higher-order spectrum (or cumulant) have appeared in 1980s. Rosenblatt (1980) proposed a consistent estimate of the transfer function of a non-Gaussian linear process by using a non-parametric higher-order spectral estimate. In Lii and Rosenblatt (1982), this estimate is used to match a non-parametric non-minimum phase MA model. Tugnait (1986) applied the basic approach of Lii and Rosenblatt (1982) to parametric non-minimum phase ARMA models. In Tugnait (1986), a spectrally equivalent minimum phase system is at first estimated using the second-order statistics of the measurements. Then, the fourth-order cumulants of the measurements are used to resolve the location of the system zeros. Giannakis and Swami (1990) proposed AR and MA parameter estimators of non-Gaussian ARMA models via linear equations by using higher-order cumulants of the observations.

Compared with these works, we address the novelty of the present paper in the following two points. First, the method of identification is based on the inner outer factorization, and we treat the identifications of outer and inner functions on an equal footing. This is conducted through the stages of model setting, parameter estimation, and order estimation. Second, parameter estimators are determined by the asymptotic efficiency. Based on this evaluation, we derive the optimal estimators of outer and inner functions, respectively, and their structures are examined from the first standpoint.

The organization of the present paper is as follows. In Section 2, we state the general setting of the problem and introduce a class of estimators of transfer function by the use of the k-th ($k \geq 2$) order spectral estimate. In Section 3, asymptotic properties of the estimators are discussed, where strong consistency and asymptotic normality are shown under several regularity conditions. Then, a Cramér-Rao type inequality for the asymptotic variances of the estimators is derived. We also give the optimal estimator which attains the lower bound of the inequality.

In Sections 4 and 5, the results obtained in Section 3 are applied to two special cases. Section 4 treats the estimation of outer transfer function or minimum phase filter. There, expected results are regained from the ones in Section 3 through simple procedures. In this way, our results are shown to include the known ones based on the second-order spectrum. Section 5 treats the estimation of inner transfer function or all-pass filter. There, we examine the structure of the optimal estimator of inner transfer function, and several similarities exist between the outer and inner cases in the optimal estimators.

In Sections 6 and 7, two separate results on the outer and inner transfer functions are combined. We propose a method of identifying non-minimum phase transfer function. Roughly speaking, the outer part of a transfer function is at first identified by the usual second-order spectral method. The inner part is then identified by the higher-order spectral method. This is conducted in two stages of the identification. One is in the parameter estimation treated in Section 6, and another is in the order estimation treated in Section 7. In the order estimation, the order of the outer part is estimated based on the established criterion, e.g. on Akaike's AIC under the Gaussian assumption. In order to estimate the order of the inner part, a criterion is proposed. This is derived by introducing a distance between inner functions, and then by following the same procedure as in Akaike's AIC (Akaike (1974)). Results are summarized in Section 8, and some discussions follow.

2. Explanations of the problem and method

In the setting introduced in Section 1, we consider the problem of identifying the transfer function $H(\lambda)$ on the basis of an n sample $\{X_0, X_1, \dots, X_{n-1}\}$. The following will be assumed throughout the present paper.

ASSUMPTION 1. (i) $\{e_t\}$ has finite moments of all orders.
(ii) The linear filter $\{h_j\}$ is 1-summable, i.e. $\sum_{j=0}^{\infty} j|h_j| < \infty$.

Note that under Assumption 1, we have

$$\sum_{u_1,\dots,u_{k-1}=-\infty}^{\infty} |u_j|\, |c(u_1,\dots,u_{k-1})| < \infty, \tag{2.1}$$

for $j = 1,\dots,k-1$, $k = 2,3,\dots$, where $c(u_1,\dots,u_{k-1})$ is the k-th order cumulant of $\{X(0), X(u_1),\dots,X(u_{k-1})\}$ (see Brillinger ((1975), Section 2.6)).

We define several quantities in the frequency domain. The k-th order cumulant spectrum of $\{X_t\}$ is given by

$$\begin{aligned} f(\lambda_1,\dots,\lambda_{k-1}) &\\ = (2\pi)^{-k+1} &\sum_{u_1,\dots,u_{k-1}=-\infty}^{\infty} c(u_1,\dots,u_{k-1}) \exp\Bigl(-i\sum_{j=1}^{k-1} u_j\lambda_j\Bigr). \end{aligned} \tag{2.2}$$

We note that $f(\lambda_1,\dots,\lambda_{k-1})$ is generally complex-valued, and it has bounded and uniformly continuous derivatives in $\lambda_1,\dots,\lambda_{k-1}$ in view of (2.1). In order to maintain the symmetry among $\lambda_1,\dots,\lambda_k$, we adopt the following notational convention. Let P_k be the set of frequency vectors satisfying $\sum_{j=1}^{k}\lambda_j = 0$,

$$P_k = \Bigl\{\lambda = (\lambda_1,\dots,\lambda_k) \in (-\pi,\pi]^k \Bigm| \sum_{j=1}^{k}\lambda_j = 0\Bigr\}. \tag{2.3}$$

Hereafter, any frequency vector λ will be assumed to belong to P_k, and we will write e.g. the k-th order cumulant spectrum as $f_k(\lambda)$. When $\{X_t\}$ is a linear process defined by (1.1), $f_k(\lambda)$ is expressed as

$$f_k(\lambda) = (2\pi)^{-k+1} K_k \prod_{j=1}^{k} H(\lambda_j), \tag{2.4}$$

where K_k denotes the k-th order cumulant spectrum of the sequence $\{e_t\}$. The k-th order cumulant spectral measure of $\{X_t\}$ is given by

$$F_k(\lambda) = \int_{-\pi}^{\lambda} f_k(\alpha)d\alpha, \tag{2.5}$$

where, the abbreviated notation in the right-hand side implies

$$\int_{-\pi}^{\lambda_1} \cdots \int_{-\pi}^{\lambda_{k-1}} f_k(\alpha_1, \ldots, \alpha_{k-1})d\alpha_1 \cdots d\alpha_{k-1}.$$

We next define several statistics. Let

$$d_n(\lambda) = \sum_{t=0}^{n-1} X_t \exp(-i\lambda t) \tag{2.6}$$

be the finite Fourier transform of the sample $\{X_0, X_1, \ldots, X_{n-1}\}$. Then, the k-th order sample periodogram is given by

$$I_k^{(n)}(\lambda) = (2\pi)^{-k+1}n^{-1}\prod_{j=1}^{k} d_n(\lambda_j). \tag{2.7}$$

Multiplying by a factor, we modify this to

$$J_k^{(n)}(\lambda) = \Psi(\lambda)I_k^{(n)}(\lambda). \tag{2.8}$$

In (2.8), $\Psi(\lambda)$ is a function satisfying $\Psi(\lambda) = 1$, if $\sum_{j=1}^{k} \lambda_j = 0$ but $\sum_{j\in J} \lambda_j \neq 0$, where J is any nonvacuous proper subset of $\{1, \ldots, k\}$, and $\Psi(\lambda) = 0$, otherwise. The factor Ψ is essential in the point of suppressing any contribution from a proper subset of P_k (see Brillinger and Rosenblatt ((1967), Section 2) and Keenan (1987)). Using the modified $J_k^{(n)}(\lambda)$, the sample k-th order cumulant spectral distribution is defined as

$$F_k^{(n)}(\lambda) = (2\pi/n)^k \sum_{s} J_k^{(n)}(2\pi s/n), \tag{2.9}$$

where, $s = (s_1, \ldots, s_k)$, and each integer s_j is summed in the range $-\pi < 2\pi s_j/n \leq \lambda_j$.

Next, we set a candidate model for linear filters. It is a model whose transfer functions or impulse responses are specified by a finite dimensional vector parameter $\theta \in \Theta$, possibly not having an ARMA representation. The parametric space Θ of θ is assumed to be a subset of an Euclidean space R^s. Our model is then expressed as

$$M = \left\{ H(\lambda, \theta) = \sum_{j=0}^{\infty} h_j(\theta)\exp(-ij\lambda) \mid \theta \in \Theta \subset R^s \right\} \tag{2.10}$$

in terms of parameterized transfer functions. We assume that the model M and its parameterization are well-suited ones as follows.

ASSUMPTION 2. (i) $H(\cdot) = H(\cdot, \theta_0)$ for a unique $\theta_0 \in \operatorname{int}\Theta$, where $\operatorname{int}\Theta$ denotes the interior of Θ.

(ii) $H(\cdot, \theta_1) = H(\cdot, \theta_2)$ if and only if $\theta_1 = \theta_2$.

The identification is then reduced to the estimation of the parameter value θ_0. In the present paper, we adopt the following type of estimators. Let y be a complex vector valued function of the form

$$y : P_k \times \Theta \to C^s, \tag{2.11}$$

i.e. $y(\lambda, \theta) = (y_1(\lambda, \theta), \ldots, y_s(\lambda, \theta))'$, and each $y_i(\lambda, \theta)$ is generally complex-valued, where $'$ denotes the transpose, so that y will be regarded as a row vector. The integration of $y(\lambda, \theta)$ with respect to the k-th order sample cumulant spectral distribution $F_k^{(n)}(\lambda)$ will be denoted by $Y_k^{(n)}(\theta)$,

$$Y_k^{(n)}(\theta) = \int_{-\pi}^{\pi} y(\lambda, \theta) dF_k^{(n)}(\lambda). \tag{2.12}$$

This is an s-dimensional vector valued function defined on Θ. The deterministic version of $Y_k^{(n)}(\theta)$ is defined replacing $F_k^{(n)}(\lambda)$ with $F_k(\lambda)$ as

$$Y_k(\theta) = \int_{-\pi}^{\pi} y(\lambda, \theta) dF_k(\lambda). \tag{2.13}$$

The estimator $\hat{\theta}_n$ associated with $Y_k^{(n)}(\theta)$ is defined, if it exists, by a solution of the simultaneous estimating equations

$$Y_k^{(n)}(\theta) = 0. \tag{2.14}$$

Similar types of estimators have been considered in estimating the time series parameters (e.g. Whittle (1953), Walker (1964), Ibragimov (1967), Taniguchi (1981), Hosoya and Taniguchi (1982), Keenan (1985)). In these works, estimators are mostly defined as a minimization solution of an integral expression. In the present paper, it is not assumed that $Y_k^{(n)}(\theta)$ or $Y_k(\theta)$ is the gradient vector of a function. Therefore, estimators which are not necessarily related to an optimization procedure are also considered.

In the next section, we state conditions under which strong consistency and asymptotic normality of the proposed type of estimators are guaranteed. Based on these results, the optimal estimator in the sense of having the minimal asymptotic variance will be derived.

3. Strong consistency and asymptotic normality

Keenan (1987) has already studied the asymptotic theory of functionals of the k-th ($k \geq 2$) order sample cumulant spectra. In the light of Keenan (1987), we give Lemmas 3.1 and 3.2, Theorems 3.1 and 3.2 below without proof.

Conditions for the consistency of estimators are the following.

CONDITION 1. (i) Any element $y_j(\lambda, \theta)$ of $y(\lambda, \theta)$ is continuous in λ and θ, which is of bounded variation in any λ_j with $\sup_{\theta \in \Theta} \| \, |y(\cdot, \theta)|_\infty \|_v < \infty$, where $| \cdot |_\infty$ is the sup norm on R^s, and $\| \cdot \|_v$ is the total variation norm.

(ii) The simultaneous equations $Y_k(\theta) = 0$ have a unique solution θ_0 whatever the value of θ_0 in Θ.

Let $\| \cdot \|_\infty$ be the sup norm on P_k,

$$\|K\|_\infty = \sup_{\lambda \in P_k} |K(\lambda)|,$$

where $K(\cdot)$ is a function on P_k.

LEMMA 3.1. *Under Assumptions* 1 *and* 2,

$$\|F_k^{(n)} - F_k\|_\infty \to 0 \qquad w.p.1.$$

THEOREM 3.1. *Under Assumptions* 1, 2 *and Condition* 1, *there exists a solution $\hat{\theta}_n$ (i.e. $Y_k^{(n)}(\theta) = 0$ have a solution at $\hat{\theta}_n$) such that $\hat{\theta}_n \to \theta_0$ with probability one.*

Conditions for the asymptotic normality of the consistent estimators are the following.

CONDITION 2. (i) Any $\partial y_i(\lambda, \theta)/\partial \theta^j$ is continuous in λ and θ, and of bounded variation in λ for all θ.

(ii) The $s \times s$ matrix W is non-singular, where $W = [w_{ij}]$ is given by

$$W = -\int_{-\pi}^{\pi} \partial y(\lambda, \theta_0)/\partial \theta' dF_k(\lambda), \tag{3.1}$$

i.e.

$$w_{ij} = -\int_{-\pi}^{\pi} \partial y_i(\lambda, \theta_0)/\partial \theta^j dF_k(\lambda).$$

(iii) $y_i(\lambda, \theta)$ is symmetric in any λ_j.

LEMMA 3.2. *Under Assumptions* 1, 2 *and Conditions* 1, 2,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n} A_n W^{-1} \left\{ O(n^{-1}) + \int_{-\pi}^{\pi} y(\lambda, \theta_0) d[F_k^{(n)}(\lambda) - F_k(\lambda)] \right\}, \tag{3.2}$$

where the components of the matrix $A_n \to 1$ w.p.1. and the $O(n^{-1})$ term does not depend on the realization.

Under the stated conditions for $y(\lambda, \theta_0)$, $\sqrt{n}[Y_k^{(n)}(\theta_0) - Y_k(\theta_0)]$ is asymptotically normal with mean zero and covariance matrix V given by

$$V = 2\pi \sum_{\underline{P}} \int_{-\pi}^{\pi} y(\alpha, \theta_0) y'(-\underline{P}\alpha, \theta_0) \prod_{j=1}^{k} f_2(\alpha_j) d\alpha,$$

where $\underline{P}$ denotes the permutation of $\alpha = (\alpha_1, \ldots, \alpha_k)$, and the sum $\sum_{\underline{P}}$ ranges over all permutations. This is the k-th order analogue of Theorem 5.10.1, Brillinger (1975) applied to $y(\lambda, \theta_0)$. In the present case, $f_2(\alpha_j) = (2\pi)^{-1} K_2 \cdot H(\alpha_j) H(-\alpha_j)$, so that

$$\prod_{j=1}^{k} f_2(\alpha_j) = (K_2/2\pi)^k S_k(\alpha) S_k(-\alpha),$$

where $S_k(\alpha) = \prod_{j=1}^{k} H(\alpha_j)$. Hence, the covariance matrix V is written in view of Condition 2(iii) as

$$V = (2\pi)^{-k+1} K_2^k k! \int_{-\pi}^{\pi} y(\lambda, \theta_0) S_k(\lambda, \theta_0) y'(-\lambda, \theta_0) S_k(-\lambda, \theta_0) d\lambda.$$

On the other hand, we have for the matrix W

$$\begin{aligned} W &= -(2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} \partial y(\lambda, \theta_0)/\partial\theta' S_k(\lambda, \theta_0) d\lambda \\ &= (2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} y(\lambda, \theta_0) \partial S_k(\lambda, \theta_0)/\partial\theta' d\lambda. \end{aligned}$$

The last expression is obtained by differentiating

$$Y_k(\theta_0) = (2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} y(\lambda, \theta_0) S_k(\lambda, \theta_0) d\lambda = 0$$

with respect to θ_0, since it is an identity in θ_0 in view of Condition 1(ii). Then, we have the following theorem by Lemma 3.2.

THEOREM 3.2. *Under Assumptions* 1, 2 *and Conditions* 1, 2, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is asymptotically normal with mean zero and covariance matrix* Ω *given by*

$$\Omega = W^{*-1} V W^{-1}, \tag{3.3}$$

where

$$W = (2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} y(\lambda, \theta_0) \partial S_k(\lambda, \theta_0)/\partial\theta' d\lambda, \tag{3.4}$$

$$V = (2\pi)^{-k+1} K_2^k k! \int_{-\pi}^{\pi} y(\lambda, \theta_0) S_k(\lambda, \theta_0) y'(-\lambda, \theta_0) S_k(-\lambda, \theta_0) d\lambda, \tag{3.5}$$

$$S_k(\lambda, \theta) = \prod_{j=1}^{k} H(\lambda_j, \theta), \tag{3.6}$$

and $W^* = [w_{ij}^*]$ *denotes the complex conjugate.*

There exists a Cremér-Rao type inequality for the asymptotic variances, and we give the optimal estimator attaining the lower bound. Let $\tilde{y}(\lambda, \theta)$ be the s-dimensional vector valued function given by

$$\tilde{y}(\lambda,\theta) = \{S_k(\lambda,\theta)S_k(-\lambda,\theta)\}^{-1}\partial S_k(-\lambda,\theta)/\partial\theta. \tag{3.7}$$

We denote by $\tilde{\theta}_n$ the estimator which is a solution of a set of estimating equations

$$\tilde{Y}_k^{(n)}(\theta) = \int_{-\pi}^{\pi} \tilde{y}(\lambda,\theta)dF_k^{(n)}(\lambda) = 0. \tag{3.8}$$

Then, the optimality result is given by the following theorem.

THEOREM 3.3. *If* $\tilde{y}(\lambda,\theta)$ *satisfy Conditions* 1 *and* 2, *the covariance matrix* Ω *of* $\hat{\theta}_n$ *is bounded by* Ω_*,

$$\Omega \geq \Omega_* = K_2^k K_k^{-2}(k-1)!G^{-1}(\theta_0), \tag{3.9}$$

where

$$G(\theta_0) = (2\pi)^{-1}\int_{-\pi}^{\pi} \partial \log H(\lambda,\theta_0)/\partial\theta \partial \log H(-\lambda,\theta_0)/\partial\theta' d\lambda \tag{3.10}$$

and the inequality implies that $\Omega - \Omega_*$ *is a positive semi-definite matrix. The lower bound* Ω_* *is attained by the estimator* $\tilde{\theta}_n$.

PROOF. Let us define

$$\alpha = \sqrt{n}[\hat{\theta}_n - \theta_0, \tilde{\theta}_n - \theta_0]' \quad \text{and} \quad A = \begin{bmatrix} W^{*-1}VW^{-1} & \Omega_0 \\ \Omega_0 & \Omega_0 \end{bmatrix},$$

where

$$\Omega_0 = K_2^k K_k^{-2} k! G_k^{-1}(\theta_0), \tag{3.11}$$

$$G_k(\theta_0) = (2\pi)^{-k+1}\int_{-\pi}^{\pi} \partial \log S_k(\lambda,\theta_0)/\partial\theta \partial \log S_k(-\lambda,\theta_0)/\partial\theta' d\lambda. \tag{3.12}$$

Then, it is easily shown that α is asymptotically normal with mean zero and covariance matrix given by A. We only explain the derivation of A. The first diagonal part is the result of Theorem 3.2, and the second diagonal part is given simply by substituting

$$\tilde{y}(\lambda,\theta_0) = \{S_k(\lambda,\theta_0)S_k(-\lambda,\theta_0)\}^{-1}\partial S_k(-\lambda,\theta_0)/\partial\theta$$

for $y(\lambda, \theta_0)$ in Theorem 3.2. The off diagonal part is derived by noting that the covariance matrix between $\int_{-\pi}^{\pi} y(\lambda, \theta_0) dF_k^{(n)}(\lambda)$ and $\int_{-\pi}^{\pi} \tilde{y}(\lambda, \theta_0) dF_k^{(n)}(\lambda)$ is

$$\begin{aligned}(2\pi)^{-k+1} K_2^k k! \int_{-\pi}^{\pi} & y(\lambda, \theta_0) S_k(\lambda, \theta_0) \tilde{y}'(-\lambda, \theta_0) S_k(-\lambda, \theta_0) d\lambda \\ &= (2\pi)^{-k+1} K_2^k k! \int_{-\pi}^{\pi} y(\lambda, \theta_0) \partial S_k(\lambda, \theta_0)/\partial\theta' d\lambda \\ &= K_2^k K_k^{-1} k! W.\end{aligned}$$

Since A is a covariance matrix, it is positive semi-definite. Hence,

$$[I \ \ -I] \begin{bmatrix} W^{*-1} V W^{-1} & \Omega_0 \\ \Omega_0 & \Omega_0 \end{bmatrix} \begin{bmatrix} I \\ -I \end{bmatrix} \geq 0.$$

Therefore, $\Omega = W^{*-1} V W^{-1} \geq \Omega_0$.

We next show that Ω_0 is equal to Ω_*. From $S_k(\lambda, \theta) = \prod_{j=1}^{k} H(\lambda_j, \theta)$, we have

$$\partial \log S_k(\lambda, \theta_0)/\partial\theta = \sum_{j=1}^{k} \partial \log H(\lambda_j, \theta_0)/\partial\theta,$$

so that

$$\begin{aligned} G_k(\theta_0) &= \sum_{i,j=1}^{k} (2\pi)^{-k+1} \int_{-\pi}^{\pi} \partial \log H(\lambda_i, \theta_0)/\partial\theta \partial \log H(-\lambda_j, \theta_0)/\partial\theta' d\lambda \\ &= k\{G(\theta_0) + (k-1)\omega(\theta_0)\omega'(\theta_0)\} \\ &= kG(\theta_0), \end{aligned}$$

where we have put

$$\omega(\theta_0) = (2\pi)^{-1} \int_{-\pi}^{\pi} \partial \log H(\lambda, \theta_0)/\partial\theta d\lambda,$$

and used the result of Lemma 3.3 below, i.e. $\omega(\theta_0) = 0$. Hence, we have

$$\Omega_0 = K_2^k K_k^{-2} k! G_k^{-1}(\theta_0) = K_2^k K_k^{-2} (k-1)! G^{-1}(\theta_0) = \Omega_*,$$

proving the theorem.

LEMMA 3.3. *If $\tilde{y}(\lambda, \theta)$ satisfy Conditions 1 and 2,*

$$\omega(\theta_0) = (2\pi)^{-1} \int_{-\pi}^{\pi} \partial \log H(\lambda, \theta_0)/\partial\theta d\lambda = 0. \tag{3.13}$$

PROOF. Let us define $\tilde{Y}_k(\theta) = \int_{-\pi}^{\pi} \tilde{y}(\lambda, \theta) dF_k(\lambda)$, which is the deterministic version of $\tilde{Y}_k^{(n)}(\theta)$. By Condition 1(ii), $\tilde{Y}_k(\theta_0) = 0$, and the left-hand side becomes

$$\begin{aligned}\tilde{Y}_k(\theta_0) &= (2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} \{S_k(\lambda, \theta_0) S_k(-\lambda, \theta_0)\}^{-1} \partial S_k(-\lambda, \theta_0)/\partial\theta S_k(\lambda, \theta_0) d\lambda \\ &= (2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} \partial \log S_k(\lambda, \theta_0)/\partial\theta d\lambda \\ &= (2\pi)^{-k+1} K_k \sum_{j=1}^{k} \int_{-\pi}^{\pi} \partial \log H(\lambda_j, \theta_0)/\partial\theta d\lambda \\ &= K_k k \omega(\theta_0).\end{aligned}$$

Hence, $\omega(\theta_0) = 0$, proving the lemma.

In the next two sections, we treat the estimations of outer and inner transfer functions, respectively. We focus on the optimal estimators, and the results of Theorem 3.3 are applied to each case.

4. Estimation of the outer transfer function

In this section, the model is assumed to be the following set of transfer functions specified by a vector parameter $\xi \in \Xi$.

$$M_O = \Big\{ A(\lambda, \xi) = \sum_{j=0}^{\infty} a_j(\xi) \exp(-ij\lambda) \mid \xi \in \Xi, a_0(\xi) \equiv 1, \\ A(z, \xi) \text{ is analytic and has no roots in } |z| \le 1 \Big\}. \tag{4.1}$$

This is the model customly used under the Gaussian assumption. We show that the known results are derived from Theorems 3.2 and 3.3 by putting $k = 2$.

When $k = 2$, we have at first for $S_2(\lambda, \xi)$ (3.6)

$$S_2(\lambda, \xi) = A(\lambda, \xi) A(-\lambda, \xi), \tag{4.2}$$

so that $S_2(\lambda, \xi) = S_2(-\lambda, \xi)$. The matrices W (3.4) and V (3.5) are

$$W = (2\pi)^{-1} K_2 \int_{-\pi}^{\pi} y(\lambda, \xi_0) \partial S_2(\lambda, \xi_0)/\partial\xi' d\lambda, \tag{4.3}$$

$$V = (2\pi)^{-1} K_2^2 2 \int_{-\pi}^{\pi} y(\lambda, \xi_0) S_2(\lambda, \xi_0) y'(-\lambda, \xi_0) S_2(\lambda, \xi_0) d\lambda, \tag{4.4}$$

respectively. From (3.9), the asymptotic variance $\Omega = W^{*-1} V W^{-1}$ is bounded by Ω_*,

$$\Omega_* = K_2^2 K_2^{-2} G^{-1}(\xi_0) = G^{-1}(\xi_0), \tag{4.5}$$

where

$$G(\xi_0) = (2\pi)^{-1} \int_{-\pi}^{\pi} \partial \log A(\lambda, \xi_0)/\partial\xi \partial \log A(-\lambda, \xi_0)/\partial\xi' d\lambda. \tag{4.6}$$

The lower bound can also be written in spectral representation, which is given by Ω_0 (3.11) with $G_2(\xi_0)$ (3.12),

$$\Omega_0 = K_2^2 K_2^{-2} 2! G_2^{-1}(\xi_0) = 2G_2^{-1}(\xi_0), \tag{4.7}$$

where

$$G_2(\xi_0) = (2\pi)^{-1} \int_{-\pi}^{\pi} \partial \log S_2(\lambda, \xi_0)/\partial\xi \partial \log S_2(\lambda, \xi_0)/\partial\xi' d\lambda. \tag{4.8}$$

From (3.7), this is attained by

$$\begin{aligned} \tilde{y}(\lambda, \xi) &= S_2^{-2}(\lambda, \xi) \partial S_2(\lambda, \xi)/\partial\xi \\ &= -\partial S_2^{-1}(\lambda, \xi)/\partial\xi. \end{aligned} \tag{4.9}$$

Neglecting the sign, the optimal $\tilde{y}(\lambda, \xi)$ is the gradient of the function $S_2^{-1}(\lambda, \xi)$. Therefore, the vector

$$\tilde{Y}_2^{(n)}(\xi) = \int_{-\pi}^{\pi} \tilde{y}(\lambda, \xi) dF_2^{(n)}(\lambda)$$

is also the gradient of

$$D^{(n)}(\xi) = \int_{-\pi}^{\pi} S_2^{-1}(\lambda, \xi) dF_2^{(n)}(\lambda), \tag{4.10}$$

where

$$F_2^{(n)}(\lambda) = (2\pi/n) \sum_s I_2^{(n)}(2\pi s/n), \qquad -\pi < 2\pi s/n \le \lambda, \tag{4.11}$$

and

$$I_2^{(n)}(\lambda) = (2\pi n)^{-1} \left| \sum_{t=0}^{n-1} X_t \exp(-it\lambda) \right|^2 \tag{4.12}$$

is the usual sample periodogram.

It is known that the deterministic version of $D^{(n)}(\xi)$

$$D(\xi) = \int_{-\pi}^{\pi} S_2^{-1}(\lambda, \xi) dF_2(\lambda) \tag{4.13}$$

satisfies the following condition.

CONDITION 1. (ii)$'$ $D(\xi) \geq 1$, and the minimum $D(\xi) = 1$ is attained at the single value ξ_0 (see e.g. Hannan (1973)).

Therefore, the optimal estimate $\tilde{\xi}_n$ corresponds to a minimizing value of $D^{(n)}(\xi)$. Note that Condition 1(ii)$'$ guarantees the consistency of $\tilde{\xi}_n$ in view of Condition 1(ii). It is also well known that $\tilde{\xi}_n$ is asymptotically equivalent to the maximum likelihood estimator under the Gaussian assumption (Hannan (1973)).

We have thus regained several known results in the estimation of minimum phase filter. A point to be noted is that the optimal estimate $\tilde{\xi}_n$ is necessarily a minimizing value of $D^{(n)}(\xi)$. There exists one more case where the optimal estimate is given by an extreme value of a function. It is the subject of the next section.

5. Estimation of the inner transfer function

In this section, the model is assumed to be the following set of transfer functions specified by a vector parameter $\eta \in H$,

$$(5.1) \qquad M_I = \Big\{ B(\lambda, \eta) = \sum_{j=0}^{\infty} b_j(\eta) \exp(-ij\lambda) \mid \eta \in H, \\ B(z,\eta) \text{ is analytic}, B(z,\eta)B(z^{-1},\eta) < 1 \text{ in } |z| < 1, \\ \text{and } B(z,\eta)B(z^{-1},\eta) = 1 \text{ on } |z| = 1 \Big\}.$$

That is, the model consists of all-pass filters or inner transfer functions. Note that by the condition $B(z)B(z^{-1}) = 1$ on $|z| = 1$, the presence of the all-pass filter has no effect on the second-order spectrum of $\{X_t\}$, so a higher-order (i.e. $k \geq 3$) spectrum will be employed for identifying the inner transfer function. Here, we assume that for the selected order k of the spectrum, the sign of K_k (the k-th order spectrum of $\{e_t\}$) is known. Then, without loss of generality, it will be hereafter assumed

ASSUMPTION 3. $K_k > 0$.

The necessity of Assumption 3 will be explained later.

We examine the optimal estimator. From (3.7),

$$(5.2) \qquad \tilde{y}(\lambda, \eta) = \{S_k(\lambda, \eta) S_k(-\lambda, \eta)\}^{-1} \partial S_k(-\lambda, \eta)/\partial \eta,$$

where

$$(5.3) \qquad S_k(\lambda, \eta) = \prod_{j=1}^{k} B(\lambda_j, \eta).$$

Since $B(\lambda, \eta)B(-\lambda, \eta) = 1$ holds also for S_k, $S_k(\lambda, \eta)S_k(-\lambda, \eta) = 1$, $\tilde{y}(\lambda, \eta)$ is simplified to

$$(5.4) \qquad \tilde{y}(\lambda, \eta) = \partial S_k(-\lambda, \eta)/\partial \eta,$$

i.e. the optimal $\tilde{y}$ is again a gradient vector. Therefore, by defining the function

$$E_k^{(n)}(\eta) = \int_{-\pi}^{\pi} S_k(-\lambda, \eta) dF_k^{(n)}(\lambda), \tag{5.5}$$

estimating equations $\tilde{Y}_k^{(n)}(\eta) = 0$ can be written as

$$\partial E_k^{(n)}(\eta)/\partial\eta = 0, \tag{5.6}$$

and the optimal estimate $\tilde{\eta}_n$ gives an extreme value of $E_k^{(n)}(\eta)$.

To investigate the character of the extreme value, we take the deterministic version $E_k(\eta)$ of $E_k^{(n)}(\eta)$,

$$\begin{aligned} E_k(\eta) &= \int_{-\pi}^{\pi} S_k(-\lambda, \eta) dF_k(\lambda) \\ &= (2\pi)^{-k+1} K_k \int_{-\pi}^{\pi} S_k(-\lambda, \eta) S_k(\lambda, \eta_0) d\lambda, \end{aligned} \tag{5.7}$$

and introduce a function space L_k^2,

$$L_k^2 = \left\{ F(\lambda) \mid (2\pi)^{-k+1} \int_{-\pi}^{\pi} F(\lambda) F(-\lambda) d\lambda < \infty \right\}. \tag{5.8}$$

L_k^2 is a Hilbert space with inner product $\langle\, , \rangle_k$ defined by

$$\langle F, G \rangle_k = (2\pi)^{-k+1} \int_{-\pi}^{\pi} F(\lambda) G(-\lambda) d\lambda, \tag{5.9}$$

and the norm $\|F\|_k$ of $F \in L_k^2$ is given by

$$\|F\|_k = \sqrt{\langle F, F \rangle_k}. \tag{5.10}$$

Using these notations, $E_k(\eta)$ can be expressed as

$$E_k(\eta) = K_k \langle S_k(\eta_0), S_k(\eta) \rangle_k, \tag{5.11}$$

where

$$\|S_k(\eta_0)\|_k = \|S_k(\eta)\|_k = 1. \tag{5.12}$$

Then, it is easily shown by Schwarz's inequality and by Assumption 3 that

$$-K_k \leq E_k(\eta) \leq K_k, \tag{5.13}$$

where

$$E_k(\eta) = K_k \quad \text{if and only if} \quad S_k(\eta) = S_k(\eta_0), \tag{5.14}$$

and

$$E_k(\eta) = -K_k \quad \text{if and only if} \quad S_k(\eta) = -S_k(\eta_0). \tag{5.15}$$

The expression (5.11) shows that $E_k(\eta)$ in effect measures the inner product between the two unit vectors $S_k(\eta_0)$ and $S_k(\eta)$ in the function space L_k^2. From this observation, the following holds under Assumptions 2 and 3.

CONDITION 1. (ii)″ The maximum $E_k(\eta) = K_k$ is attained at the single value η_0.

This is the counterpart of Condition 1(ii)′ in the outer transfer function case, and Condition 1(ii)″ guarantees the consistency of $\tilde{\eta}_n$.

Including the representations of the lower bound for the asymptotic variances, we give the obtained results as a theorem.

THEOREM 5.1. *The optimal estimator $\tilde{\eta}_n$ on the inner transfer function model* (5.1) *is given by a maximizing value of $E_k^{(n)}(\eta)$* (5.5). *The covariance matrix of $\tilde{\eta}_n$ is given from* (3.9) *and* (3.10) *by*

$$\Omega_* = K_2^k K_k^{-2}(k-1)!G^{-1}(\eta_0), \tag{5.16}$$

where

$$G(\eta_0) = (2\pi)^{-1}\int_{-\pi}^{\pi} \partial B(\lambda, \eta_0)/\partial\eta \partial B(-\lambda, \eta_0)/\partial\eta' d\lambda, \tag{5.17}$$

or from (3.11) *and* (3.12) *by*

$$\Omega_0 = K_2^k K_k^{-2} k! G_k^{-1}(\eta_0), \tag{5.18}$$

where

$$G_k(\eta_0) = (2\pi)^{-k+1}\int_{-\pi}^{\pi} \partial S_k(\lambda, \eta_0)/\partial\eta \partial S_k(-\lambda, \eta_0)/\partial\eta' d\lambda. \tag{5.19}$$

A typical finite dimensional model is an ARMA one, i.e. a set of transfer functions which are rational functions of $z = \exp(-i\lambda)$. In the case of outer transfer function, it takes the form

$$\begin{aligned} M_O(p,q) = \Big\{ & A(z,\xi) = g(z)/f(z) \mid \\ & f(z) = 1 + a_1 z + \cdots + a_p z^p, g(z) = 1 + b_1 z + \cdots + b_q z^q, \\ & f(z) \text{ and } g(z) \text{ have no roots in } |z| \le 1 \Big\}, \end{aligned} \tag{5.20}$$

where, the $(p+q)$-dimensional vector $\xi = (a_1, \dots, a_p, b_1, \dots, b_q)$ is the set of parameters specifying the transfer functions in $M_O(p,q)$. On the other hand, the model of rational inner transfer functions takes the form

$$\begin{aligned} M_I(r) = \{ & B(z,\eta) = g(z)/f(z) \mid \\ & f(z) = 1 + c_1 z + \cdots + c_r z^r, \\ & g(z) = z^r f(z^{-1}) = c_r + c_{r-1} z + \cdots + z^r, \\ & f(z) \text{ has no roots in } |z| \le 1\}, \end{aligned} \tag{5.21}$$

where, the r-dimensional vector $\eta = (c_1, \ldots, c_r)$ is the set of parameters specifying the transfer functions in $M_I(r)$. Note that by the condition $B(z)B(z^{-1}) = 1$ on $|z| = 1$, the numerator $g(z)$ is uniquely determined as the reciprocal polynomial of the denominator $f(z)$.

As for the ARMA model $M_O(p, q)$, the following was shown by Åström and Söderström (1974). Let (p_0, q_0) be the true orders of $A(z, \xi_0)$. Then, there is a unique local and global minimum of the function $D(\xi)$ on the model $M_O(p_0, q_0)$. It means that

$$D(\xi) \geq 1, \qquad D(\xi) = 1 \quad \text{if and only if} \quad \xi = \xi_0, \tag{5.22}$$

and

$$\operatorname{grad} D(\xi) = 0 \quad \text{if and only if} \quad \xi = \xi_0 \tag{5.23}$$

on $M_O(p_0, q_0)$. This result is useful when seeking the optimal estimate $\tilde{\xi}_n$ based on the gradient method.

We consider the same problem for the function $E_k(\eta)$ defined on the rational inner transfer function model $M_I(r)$. The result is immediately obtained in view of the expression (5.11) for $E_k(\eta)$.

THEOREM 5.2. *Let r_0 be the true order of $B(z, \eta_0)$. Then, there is a unique local and global maximum of $E_k(\eta)$ on $M_I(r_0)$. That is,*

$$E_k(\eta) \leq K_k, \qquad E_k(\eta) = K_k \quad \text{if and only if} \quad \eta = \eta_0, \tag{5.24}$$

and

$$\operatorname{grad} E_k(\eta) = 0 \quad \textit{if and only if} \quad \eta = \eta_0 \tag{5.25}$$

on $M_I(r_0)$.

PROOF. (5.24) is already given by Condition 1(ii)$''$. We take a unit sphere $\|F\|_k = 1$ in L_k^2, and assume that the true vector $S_k(\eta_0)$ points to the north pole. The model $M_I(r_0)$ is an r_0-dimensional region in the sphere containing the north pole. Note that by the definition, the south pole is excluded from $M_I(r_0)$. Then, (5.25) holds, since the longitudinal components of $\operatorname{grad} E_k(\eta)$ do not vanish at $\eta \neq \eta_0$.

Theorems 5.1 and 5.2 show that there exist similarities between the outer and inner cases in the structures of optimal estimators.

6. Estimation of non-minimum phase transfer function

In this section, we consider the problem of estimating a non-minimum phase transfer function. The results obtained in Sections 4 and 5 will be used as building blocks here.

We begin by setting a reasonable model for the non-minimum phase transfer functions. The phrase "reasonable" implies that Condition 1(ii) holds for the optimal $\tilde{y}(\lambda, \theta)$ (3.7) on the adopted model. If Condition 1(ii) does not hold on a model, the consistency of the optimal estimator is not guaranteed there. As a simple example, let us take a non-minimum phase MA(1) model which is defined as

$$\text{MA}(1) = \{H(z,\theta) = 1 + \theta z \mid \theta \in R^1\}. \tag{6.1}$$

Under Condition 1(ii), θ_0 must be a solution of $\tilde{Y}_k(\theta) = 0$, i.e.

$$\tilde{Y}_k(\theta_0) = 0. \tag{6.2}$$

The left-hand side is proportional to $\omega(\theta_0)$ given by (3.13) (see the proof of Lemma 3.3), so that the condition (6.2) is

$$\omega(\theta_0) = (2\pi i)^{-1} \oint_{|z|=1} \partial \log H(z, \theta_0)/\partial\theta z^{-1} dz = 0. \tag{6.3}$$

We have for $H(z,\theta) \in \text{MA}(1)$,

$$\partial \log H(z, \theta_0)/\partial\theta = z/(1 + \theta_0 z).$$

Then,

$$\omega(\theta_0) = (2\pi i)^{-1} \oint_{|z|=1} (1 + \theta_0 z)^{-1} dz,$$

which is not zero if $|\theta_0| > 1$. Hence, Condition 1(ii) does not hold for the optimal estimator θ_n on MA(1), and so MA(1) is not a reasonable model. In general, the non-minimum phase ARMA model $M(p,q)$ defined simply by dropping the invertibility condition as

$$\begin{aligned} M(p,q) = \{H(z,\theta) = g(z)/f(z) \mid \\ f(z) = 1 + a_1 z + \cdots + a_p z^p, g(z) = 1 + b_1 z + \cdots + b_q z^q, \\ f(z) \text{ has no roots in } |z| \le 1\} \end{aligned} \tag{6.4}$$

is not a reasonable model.

Let us recall the conditions satisfied by the two optimal estimators, i.e. Condition 1(ii)$'$ in Section 4 and Condition 1(ii)$''$ in Section 5. We noted that Conditions 1(ii)$'$ and 1(ii)$''$ guarantee the consistency of the optimal estimators in respective cases. Based on this fact and the inner outer factorization, we introduce a non-minimum phase model M_{OI} whose parameter space Θ is the direct product of Ξ in M_O and H in M_I.

$$M_{OI} = \{H(z,\theta) = A(z,\xi)B(z,\eta) \mid A(z,\xi) \in M_O, B(z,\eta) \in M_I, \\ \text{i.e. } \theta = (\xi,\eta) \text{ or } \Theta = \Xi \times H\}. \tag{6.5}$$

On the model M_{OI}, we propose a method of estimating $H(\lambda) = H(\lambda,\theta_0)$. At first, the outer part $A(\lambda)$ of $H(\lambda)$ is estimated on the model M_O parameterized by $\xi \in \Xi$. The estimator $\tilde{\xi}_n$ is given by a minimizing value of $D^{(n)}(\xi)$ (4.10). Next, the inner part $B(\lambda)$ of $H(\lambda)$ is estimated on the model M_I parameterized by $\eta \in H$. The estimator $\bar{\eta}_n$ is given by a maximizing value of

$$\bar{E}_k^{(n)}(\eta) = \int_{-\pi}^{\pi} S_k(-\lambda,\eta) S_k^{-1}(\lambda,\tilde{\xi}_n) dF_k^{(n)}(\lambda), \tag{6.6}$$

where

$$S_k(\lambda,\tilde{\xi}_n) = \prod_{j=1}^{k} A(\lambda_j,\tilde{\xi}_n), \tag{6.7}$$

and

$$S_k(-\lambda,\eta) = \prod_{j=1}^{k} B(-\lambda_j,\eta). \tag{6.8}$$

Note that $\bar{E}_k^{(n)}(\eta)$ is slightly different from $E_k^{(n)}(\eta)$ given by (5.5). However, the strong consistency $\tilde{\xi}_n \to \xi_0$ w.p.1. erases the difference asymptotically, so that

$$\bar{E}_k^{(n)}(\eta) \to E_k(\eta), \qquad n \to \infty \tag{6.9}$$

holds uniformly in η, where $E_k(\eta)$ is the deterministic function given by (5.7). Then, by the same reasoning as in the case of Theorem 3.1, the estimator $\bar{\eta}_n$ is also strongly consistent,

$$\bar{\eta}_n \to \eta_0 \quad \text{w.p.1.} \tag{6.10}$$

This is the procedure of parameter estimations on the model M_{OI}.

7. Estimation of the orders of transfer function

We introduced ARMA models by $M_O(p,q)$ (5.20) for the outer transfer functions and by $M_I(r)$ (5.21) for the inner transfer functions, respectively. Then, we showed that under the knowledge of the true orders (p_0,q_0) and r_0, the function $D(\xi)$ has a unique local and global minimum at ξ_0 (see (5.22), (5.23)), and the function $E_k(\eta)$ has a unique local and global maximum at η_0 (see (5.24), (5.25)). Therefore, estimating these orders are crucial points in the identification.

There have been numerous works also on this problem, where estimations are made based on respective criteria. The most famous one is Akaike's AIC under the Gaussian distribution. In the frequency domain expression, it takes the form

$$\text{AIC}(s) = \log D_s^{(n)}(\tilde{\xi}_n) + 2s/n, \tag{7.1}$$

where $D_s^{(n)}(\tilde{\xi}_n)$ denotes the value of $D^{(n)}(\xi)$ at the optimal estimate $\tilde{\xi}_n$ on the s-dimensional outer transfer function model $M_O(s)$. The estimates of the order s will be obtained by the minimizing one of AIC(s) or of any other criterion (see e.g. Hannan (1980) and the references therein). We propose a criterion for estimating the order of the inner part. It is obtained by almost duplicating the procedure in deriving Akaike's AIC.

We at first treat the case where the transfer function itself is an inner function. Then, the order will be selected in the family of inner transfer function models $\{M_I(r); r = 1, 2, 3, \ldots\}$, where r denotes the dimension of the model. In the function space L_k^2 introduced by (5.8), a natural distance is defined by using the norm $\|\cdot\|_k$ (5.10). Multiplying by a constant, we define a new distance $I(\eta_1, \eta_2)$ between $S_k(\lambda, \eta_1)$ and $S_k(\lambda, \eta_2)$ by

$$I(\eta_1, \eta_2) = K_k^2 (2K_2^k k!)^{-1} \|S_k(\eta_1) - S_k(\eta_2)\|_k^2. \tag{7.2}$$

When η is sufficiently close to the true value η_0, $I(\eta_0, \eta)$ admits an approximation

$$I(\eta_0, \eta_0 + d\eta) = d\eta' \Omega_0^{-1} d\eta/2, \tag{7.3}$$

where Ω_0 is the lower bound for the asymptotic variances given by (5.18). We note that $I(\eta_0, \eta)$ can also be written as

$$\begin{aligned} I(\eta_0, \eta) &= K_k^2 (K_2^k k!)^{-1} \{1 - \langle S_k(\eta_0), S_k(\eta) \rangle_k\} \\ &= K_k (K_2^k k!)^{-1} \{K_k - E_k(\eta)\}, \end{aligned} \tag{7.4}$$

where $E_k(\eta)$ is given by (5.7) or (5.11). We take $I(\eta_0, \tilde{\eta}_n)$ as a criterion for defining a best fitting model by its minimization. That is, $I(\eta_0, \tilde{\eta}_n)$ will carry the same role as the Kullback-Leibler information played in deriving AIC (see Akaike ((1974), Section 4)).

Consider the situation where the variation of η for maximizing $E_k^{(n)}(\eta)$ is restricted to a lower dimensional subspace H of η which does not include η_0. We define η_* in H as the minimizing value of $I(\eta_0, \eta)$, i.e.

$$\min_{\eta \in H} I(\eta_0, \eta) = I(\eta_0, \eta_*), \tag{7.5}$$

and then the Pythagorean theorem

$$I(\eta_0, \tilde{\eta}_n) = I(\eta_0, \eta_*) + I(\eta_*, \tilde{\eta}_n) \tag{7.6}$$

holds. In the right-hand side, we at first examine the term $I(\eta_*, \tilde{\eta}_n)$. For the optimal estimate $\tilde{\eta}_n$ of η_0 restricted in H, if η_* is sufficiently close to η_0, it can be shown that the distribution of $2nI(\eta_*, \tilde{\eta}_n)$ for sufficiently large n is approximated under certain regularity conditions by a chi-square distribution, with the degrees of freedom equal to the dimension of the restricted parameter space. Thus, it holds that

$$E_\infty 2nI(\eta_0, \tilde{\eta}_n) = 2nI(\eta_0, \eta_*) + r, \tag{7.7}$$

where E_∞ denotes the mean of the approximate distribution and r is the dimension of H.

Next, we need to develop some estimate of $2nI(\eta_0, \eta_*)$. From the alternate expression (7.4), we have

$$I(\eta_0, \eta_*) = K_k^2(K_2^k k!)^{-1} - K_k(K_2^k k!)^{-1}E_k(\eta_*), \tag{7.8}$$

and we must seek estimates of K_k, K_2 and $E_k(\eta_*)$.

As for the variance K_2 of $\{e_t\}$, $f_2(\lambda) = K_2/2\pi$ holds, since the filter itself is all-pass. On the other hand,

$$\int_{-\pi}^{\pi} I_2^{(n)}(\lambda)d\lambda \to \int_{-\pi}^{\pi} f_2(\lambda)d\lambda = K_2 \qquad \text{w.p.1.}$$

Thus, we take as an estimate $\hat{K}_2$

$$\hat{K}_2 = \int_{-\pi}^{\pi} I_2^{(n)}(\lambda)d\lambda. \tag{7.9}$$

As for the k-th order cumulant K_k of $\{e_t\}$,

$$f_k(\lambda, \eta_0) = (2\pi)^{-k+1}K_kS_k(\lambda, \eta_0),$$

so that

$$\|f_k(\eta_0)\|_k^2 = (2\pi)^{-2(k-1)}K_k^2.$$

On the other hand,

$$\|J_k^{(n)}\|_k^2 \to \|f_k(\eta_0)\|_k^2 \qquad \text{w.p.1.}$$

Thus, we take as an estimate $\hat{K}_k^2$

$$\hat{K}_k^2 = (2\pi)^{2(k-1)}\|J_k^{(n)}\|_k^2. \tag{7.10}$$

Under Assumption 3, an estimate $\hat{K}_k$ is also given by

$$\hat{K}_k = (2\pi)^{k-1}\|J_k^{(n)}\|_k. \tag{7.11}$$

Finally, we replace $E_k(\eta_*)$ with $E_k^{(n)}(\tilde{\eta}_n)$. Then, we have as an estimate of $2nI(\eta_0, \eta_*)$

$$2n\hat{K}_k^2(\hat{K}_2^k k!)^{-1} - 2n\hat{K}_k(\hat{K}_2^k k!)^{-1}E_k^{(n)}(\tilde{\eta}_n). \tag{7.12}$$

By the same reasoning as in AIC, the last replacement causes a downward bias. This is corrected by adding r to (7.12). Furthermore, for the purpose of comparison of the values of the estimates of $E\ I(\eta_0, \tilde{\eta}_n)$ for various models, the first term in (7.12) is not necessary.

Based on these observations, an information criterion IIC(r) of η is defined as

$$\text{IIC}(r) = -2\hat{K}_k(\hat{K}_2^k k!)^{-1} E_k^{(n)}(\tilde{\eta}_n) + 2r/n. \tag{7.13}$$

The estimate of the order r will be obtained by the minimizing one of IIC(r).

We are now back to the problem of estimating the order of the inner part of a transfer function. Then, a minor modification to IIC(r) is necessary. We must replace $E_k^{(n)}(\tilde{\eta}_n)$ with $\bar{E}_k^{(n)}(\bar{\eta}_n)$ given by (6.6), and define $\overline{\text{IIC}}(r)$ as

$$\overline{\text{IIC}}(r) = -2\hat{K}_k(\hat{K}_2^k k!)^{-1} \bar{E}_k^{(n)}(\bar{\eta}_n) + 2r/n. \tag{7.14}$$

The orders of a transfer function are estimated in the following way. At first, the order s of the outer part is estimated based on AIC(s) (7.1) or on any other criterion. Next, the order r of the inner part is estimated based on the criterion $\overline{\text{IIC}}(r)$ (7.14).

The criteria IIC (7.13) and $\overline{\text{IIC}}$ (7.14) may need modifications. The proposed ones are based on the principle of parsimony that lies in Akaike's AIC.

8. Conclusions

The results obtained in the present paper can be summarized as follows.

The model of transfer functions specified by a vector parameter was introduced. By using the k-th ($k \geq 2$) order spectral estimate from the observed process, the parameter estimator was defined by a solution of the set of estimating equations.

The strong consistency and asymptotic normality of the estimators were shown under several regularity conditions. The Cramér-Rao type inequality for the asymptotic variances and the optimal estimator attaining the lower bound were obtained.

These results were applied to the estimations of outer (minimum phase) and inner (all-pass) transfer functions. There, the optimal estimates were shown to be given by the minimizing and maximizing values of two functions.

Using these optimal estimators successively, we proposed the method of identification of a non-minimum phase transfer function. In the order estimations, conventional methods estimate the order of the outer part of a transfer function. One criterion was proposed for estimating the order of the inner part. By introducing the distance between two inner functions, it was obtained following the same procedure as in Akaike's AIC.

A more rigorous mathematical analysis will be necessary to determine if the proposed methods are applicable. It will be the subject of our future work.

Acknowledgement

The author would like to express his sincere thanks to Prof. Shun-ichi Amari for his useful discussions and suggestions during the course of the present study.

References

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **19**, 716–723.

Åström, K. J. and Söderström, T. (1974). Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model, *IEEE Trans. Automat. Control*, **19**, 769–773.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*, Holt, New York.

Brillinger, D. R. and Rosenblatt, M. (1967). Asymptotic theory of estimates of k-th order spectra, *Spectral Analysis of Time Series* (ed. B. Harris), 153–188, Wiley, New York.

Duren, P. L. (1970). *Theory of H^p Spaces*, Academic Press, New York.

Giannakis, G. B. and Swami, A. (1990). On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes, *IEEE Trans. Acoust. Speech Signal Process.*, **38**, 478–495.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models, *J. Appl. Probab.*, **10**, 130–145.

Hannan, E. J. (1980). The estimation of the order of an ARMA process, *Ann. Statist.*, **8**, 1071–1081.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes, *Ann. Statist.*, **10**, 132–153.

Ibragimov, I. A. (1967). On maximum likelihood estimation of parameters of the spectral density of stationary time series, *Theory Probab. Appl.*, **120**, 115–119.

Kabaila, P. (1980). An optimality property of the least-squares estimate of the parameter of the spectrum of a purely nondeterministic time series, *Ann. Statist.*, **8**, 1082–1092.

Keenan, D. M. (1985). Asymptotic properties of minimization estimators for time series parameters, *Ann. Statist.*, **13**, 369–382.

Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra, *Ann. Statist.*, **15**, 134–151.

Lii, K. S. and Rosenblatt, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes, *Ann. Statist.*, **10**, 1195–1208.

Rosenblatt, M. (1980). Linear processes and bispectra, *J. Appl. Probab.*, **17**, 265–270.

Taniguchi, M. (1981). An estimation procedure of parameters of a certain spectral density model, *J. Roy. Statist. Soc. Ser. B*, **43**, 34–40.

Tugnait, J. K. (1986). Identification of nonminimum phase linear stochastic systems, *Automatica J. IFAC*, **22**, 457–464.

Walker, A. M. (1964). Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary nondeterministic time series, *J. Austral. Math. Soc.*, **4**, 363–384.

Whittle, P. (1953). Estimation and information in stationary time series, *Ark. Math. Astr. Fys.*, **2**, 423–434.