Uniqueness of the Maximum Likelihood Estimates of the Parameters of an ARMA Model

Åström, Karl Johan; Söderström, Torsten

Published in:
IEEE Transactions on Automatic Control

1974

Document Version:
Publisher's PDF, also known as Version of record

Uniqueness of the Maximum Likelihood Estimates of the Parameters of an ARMA Model

KARL J. ÅSTRÖM, MEMBER, IEEE, AND TORSTEN SÖDERSTRÖM

Abstract—Estimation of the parameters in a mixed autoregressive moving average process leads to a nonlinear optimization problem. The negative logarithm of the likelihood function, suitably normalized, converges to a deterministic function as the sample length increases. The local and global extrema of this function are investigated. Conditions for the existence of a unique global and local minimum are given.

I. Introduction

Maximum likelihood estimation of the parameters of an autoregressive moving average (ARMA) process was analyzed in [2]. Conditions for consistency of the maximum likelihood estimates are given in [2] and asymptotic normality and asymptotic efficiency are proven. Methods to compute the likelihood function, its gradient and its Hessian were also given [2], as well as a modified Newton–Raphson algorithm for maximizing the likelihood function. The proof of consistency of the estimates given in [2] depends crucially on the fact that the absolute maximum of the likelihood function is found at each sample length. When using any optimization algorithm based on gradients, there is always the possibility that the algorithm may converge to a local maximum. When attempting to fit ARMA models to EEG data, Bohlin [3] has also found cases where the likelihood function can conceivably have local maxima.

Since the value of the likelihood function for fixed parameters is a random variable, it is difficult to analyze the possibility of local minima. It can, however, be shown that the likelihood function, suitably normalized, will converge to a deterministic function. The purpose of this paper is to analyze the local and global extrema of this limiting function. The main result is that the local and global extrema coincide if enough parameters are used in the fitted model. It is also shown that if the number of parameters in the fitted model is sufficiently large and if the orders of either the autoregression or the moving average are correct then there is a unique local and global maximum. However, if the orders of both the autoregressive and the moving average part of the fitted model are too high then there will be many maxima. All maxima are on a manifold with the property that there is a common factor in the polynomials associated with the moving average and the autoregressive parts. If the number of parameters is not sufficiently large, then the likelihood function can have several local maxima.

Manuscript received January 25, 1974; revised June 26, 1974. This work was supported in part by the Swedish Board for Technical Development under Contract 733546.

The authors are with the Division of Automatic Control, Lund Institute of Technology, Lund, Sweden.

II. Preliminaries

In this section a precise mathematical formulation of the problems will be given. Some preliminary mathematical results to be used in the following section are also developed. Let $\{y(t),\ t = 1,2,\cdots\}$ be a stationary Gaussian stochastic process with rational spectral density. It follows from the representation theorem (see, e.g., [1]) that the process can be represented as a mixed autoregressive moving average process, i.e.,

$$A^*(q^{-1})y(t) = C^*(q^{-1})e(t) \tag{2.1}$$

where $e(t)$ is a sequence of independent normal (0,1) random variables, q^{-1} denotes the backward shift operator and the polynomials $A^*(z)$ and $C^*(z)$ are given by

$$\begin{aligned} A^*(z) &= 1 + a_1 z + \cdots + a_{n_a} z^{n_a} \\ C^*(z) &= 1 + c_1 z + \cdots + c_{n_c} z^{n_c}. \end{aligned} \tag{2.2}$$

The representation (2.1) can be uniquely chosen so that $A^*(z)$ and $C^*(z)$ do not have common factors and a_{n_a} as well as c_{n_c} are nonzero. The reciprocal polynomials $A(z)$ and $C(z)$ given by

$$A(z) = z^{n_a}A^*(z^{-1}) \qquad C(z) = z^{n_c}C^*(z^{-1}) \tag{2.3}$$

will also be used in the analysis.

Since the process $\{y(t)\}$ is stationary, the polynomial $A(z)$ must have all its zeros inside the unit circle. The polynomial $C(z)$ can be chosen in such a way that all its zeros are inside or on the unit circle. It will be assumed that $C(z)$ has no zeros on the unit circle.

The estimation of the parameters $a_1,\cdots,a_{n_a},c_1,\cdots,c_{n_c}$ with the maximum likelihood method leads to the problem of minimizing the function

$$V^N(\hat{a},\hat{c}) = V^N(\hat{a}_1,\cdots,\hat{a}_{\hat{n}_a},\hat{c}_1,\cdots,\hat{c}_{\hat{n}_c}) = \frac{1}{2N}\sum_{t=1}^{N}\epsilon^2(t) \tag{2.4}$$

see [2]. The integers $\hat{n}_a$ and $\hat{n}_c$ are given guesses or estimates of n_a and n_c. The residual $\epsilon(t)$ is a function of the observations $y(1),y(2),\cdots,y(t)$, defined by

$$\hat{C}^*(q^{-1})\epsilon(t) = \hat{A}^*(q^{-1})y(t) \tag{2.5}$$

where

$$\begin{aligned} \hat{A}^*(z) &= 1 + \hat{a}_1 z + \cdots + \hat{a}_{\hat{n}_a} z^{\hat{n}_a} \\ \hat{C}^*(z) &= 1 + \hat{c}_1 z + \cdots + \hat{c}_{\hat{n}_c} z^{\hat{n}_c}. \end{aligned} \tag{2.6}$$

The reciprocal polynomials $\hat{A}(z)$ and $\hat{C}(z)$ are defined analogously to (2.3). The polynomial $\hat{C}(z)$ can always be assumed to have all zeros inside or on the unit circle. It is also assumed here that all zeros are inside the unit circle.

Since only asymptotic properties will be considered, the initial values of (2.5) are thus not important. They can, e.g., be selected as zero. The assumption of stability means that the domain of V^N is taken as $R^{\hat{n}_a} \times H(\hat{C})$, where $H(\hat{C})$ is the subset of $R^{\hat{n}_c}$ such that the polynomial $\hat{C}(z)$ has all zeros strictly inside the unit circle. With use of (2.1) and (2.5) the residual can be written as

$$\epsilon(t) = \frac{\hat{A}^*(q^{-1})}{\hat{C}^*(q^{-1})} \frac{C^*(q^{-1})}{A^*(q^{-1})} e(t). \tag{2.7}$$

The maximum likelihood estimates of the model parameters are obtained by finding the absolute minimum of V^N. The existence of more than one local minimum may lead to wrong estimates and cause difficulties in the computations.

Since the values of V^N for fixed $\hat{a}$ and $\hat{c}$ is a random variable, it is in general very difficult to analyze the existence of possible local minima. It follows from ergodic theory, see [4], that V^N under the given assumptions converges with probability one to the function V, defined by

$$\begin{aligned} 2V(\hat{a},\hat{c}) &= \lim_{N\to\infty} 2V^N(\hat{a},\hat{c}) = E\epsilon^2(t) \\ &= \frac{1}{2\pi i} \oint \frac{\hat{A}^*(z)C^*(z)}{A^*(z)\hat{C}^*(z)} \frac{\hat{A}^*(z^{-1})C^*(z^{-1})}{A^*(z^{-1})\hat{C}^*(z^{-1})} \frac{dz}{z} \end{aligned} \tag{2.8}$$

where the integration path is the unit circle.

Using the reciprocal polynomials the function $V(\hat{a},\hat{c})$ is easily rewritten as

$$2V(\hat{a},\hat{c}) = \frac{1}{2\pi i} \oint \frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z) \cdot \frac{\hat{A}C}{A\hat{C}}(z)\, z^k \frac{dz}{z} \tag{2.9}$$

where the integer k is given by

$$k = n_a + \hat{n}_c - \hat{n}_a - n_c. \tag{2.10}$$

Note that since V is defined as the limit of V^N, V is only defined in the subset $R^{\hat{n}_a} \times H(\hat{C})$, although the integral representation (2.9) has sense for all $\hat{C}(z)$ without zeros *on* the unit circle.

The residual $\epsilon(t)$ is in fact an estimate of $e(t)$ and a perfect model satisfies

$$\frac{\hat{C}^*(q^{-1})}{\hat{A}^*(q^{-1})} = \frac{C^*(q^{-1})}{A^*(q^{-1})} \tag{2.11}$$

which requires that the integers $\hat{n}_a$ and $\hat{n}_c$ are chosen so that

$$\hat{n}_a \geqslant n_a \qquad \hat{n}_c \geqslant n_c. \tag{2.12}$$

The conditions (2.12) will appear later on in the analysis.

III. Global Minima

The global minima of the loss function V (2.8) will now be analyzed. Since the set $H(\hat{C})$ is open, the minimum may conceivably not exist. The function V will, however, become infinite on the boundary of $H(\hat{C})$ and no difficulties arise.

Theorem 1: Consider the mapping $V: R^{\hat{n}_a} \times H(\hat{C}) \to R$ defined by (2.8), then

$$2V(\hat{a},\hat{c}) \geqslant 1 \tag{3.1}$$

where equality is obtained if and only if

$$\hat{A}^*(z)C^*(z) = A^*(z)\hat{C}^*(z). \qquad \square \tag{3.2}$$

Proof: Consider the following inequality

$$\frac{1}{2\pi i} \oint \left[\frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z) - 1\right]\left[\frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z^{-1}) - 1\right] \frac{dz}{z} \geqslant 0 \tag{3.3}$$

where the integration path is the unit circle. The integral is nonnegative because the integrand is nonnegative.

Expanding the left-hand side of (3.3)

$$I_1 - I_2 - I_3 + I_4 \geqslant 0 \tag{3.4}$$

where

$$I_1 = \frac{1}{2\pi i} \oint \frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z) \frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z^{-1}) \frac{dz}{z} = 2V(\hat{a},\hat{c}) \tag{3.5}$$

$$I_2 = \frac{1}{2\pi i} \oint \frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z) \frac{dz}{z} = 1 \tag{3.6}$$

$$I_3 = \frac{1}{2\pi i} \oint \frac{\hat{A}^*C^*}{A^*\hat{C}^*}(z^{-1}) \frac{dz}{z} = I_2 = 1 \tag{3.7}$$

$$I_4 = \frac{1}{2\pi i} \oint \frac{dz}{z} = 1. \tag{3.8}$$

The integral I_2 equals 1 because $z = 0$ is the only pole of the integrand and its residue in $z = 0$ is 1. Inserting (3.5)–(3.8) in (3.4), (3.1) is obtained. Moreover, equality in (3.1) is obtained if and only if the left-hand side of (3.3) is zero. This is the case when integrand vanishes, i.e., when (3.2) is satisfied.

Remark: By equating the coefficients at equal powers of z in (3.2) a system of linear equations for the parameters of the polynomials $\hat{A}(z)$ and $\hat{C}(z)$ is obtained. Since there are $\hat{n}_a + \hat{n}_c$ variables and $\max(n_a + \hat{n}_c, \hat{n}_a + n_c)$ equations, we find that the equations have a solution only if the number of unknowns is greater than or equal to the number of parameters, i.e.,

$$\hat{n}_a + \hat{n}_c \geqslant \max(n_a + \hat{n}_c, \hat{n}_a + n_c)$$

or rewritten

$$\min(\hat{n}_a - n_a, \hat{n}_c - n_c) \geqslant 0 \tag{3.9}$$

which is nothing but (2.12). Equation (3.2) thus has a solution only if $\hat{n}_a \geqslant n_a$ and $\hat{n}_c \geqslant n_c$. Since the polynomials $A(z)$ and $C(z)$ have no common factors, we find that $A(z)$ must be a factor of $\hat{A}(z)$ and $C(z)$ a factor of $\hat{C}(z)$.

To analyze the consequences of (3.2) two different cases will be separated.

Case 1: If $\min(\hat{n}_a - n_a, \hat{n}_c - n_c) = 0$, then (3.2) has a unique solution which is given by

$$\hat{A}^*(z) \equiv A^*(z),\ \hat{C}^*(z) \equiv C^*(z). \qquad (3.10)$$

Case 2: If $\min(\hat{n}_a - n_a, \hat{n}_c - n_c) > 0$, then there are infinitely many solutions of (3.2). They can be represented by

$$\hat{A}^*(z) \equiv A^*(z)D^*(z) \qquad \hat{C}^*(z) \equiv C^*(z)D^*(z) \qquad (3.11)$$

where

$$D^*(z) = 1 + d_1 z + \cdots + d_{n_d} z^{n_d}.$$

The integer n_d equals $\min(\hat{n}_a - n_a, \hat{n}_c - n_c)$, and the coefficients $d_1, \cdots, d_{n_d}$ are arbitrary.

IV. Local Extrema

Having characterized the global minima of the loss function, the local extrema will now be analyzed, i.e., the points where the gradient of the loss function will vanish. It follows from (2.9) that the components of the gradient of V are given by

$$\frac{\partial V}{\partial \hat{a}_j} = \frac{1}{2\pi i} \oint \frac{\hat{A}C}{A\hat{C}}(z)\, \frac{C^*}{A^*\hat{C}^*}(z)\, z^{j+k-1} dz, \qquad j = 1, \cdots, \hat{n}_a$$

$$\frac{\partial V}{\partial \hat{c}_j} = -\frac{1}{2\pi i} \oint \frac{\hat{A}C}{A\hat{C}}(z)\, \frac{\hat{A}^*C^*}{A^*\hat{C}^*\hat{C}^*}(z)\, z^{j+k-1} dz, \qquad j = 1, \cdots, \hat{n}_c. \qquad (4.1)$$

The conditions for the gradient to vanish will now be investigated. Conceivably the gradient may vanish for parameter values such that $\hat{A}(z)$ and $\hat{C}(z)$ have common factors. For this purpose assume that they have exactly $\bar{n}_d$ common factors. Note that the value of $\bar{n}_d$ depends on the considered point in $R^{\hat{n}_a} \times H(\hat{C})$. Introduce the polynomials $\bar{A}^*(z)$, $\bar{C}^*(z)$, and $\bar{D}^*(z)$ of degrees $\bar{n}_a \mathrel{\hat{=}} \hat{n}_a - \bar{n}_d$, $\bar{n}_c \mathrel{\hat{=}} \hat{n}_c - \bar{n}_d$, and $\bar{n}_d$, respectively, through

$$\begin{aligned} \hat{A}^*(z) &= \bar{A}^*(z)\bar{D}^*(z) \\ \hat{C}^*(z) &= \bar{C}^*(z)\bar{D}^*(z) \end{aligned} \qquad (4.2)$$

where

$$\bar{D}^*(z) = 1 + \bar{d}_1 z + \cdots + \bar{d}_{\bar{n}_d} z^{\bar{n}_d}, \qquad (\bar{D}^*(z) = 1 \ \text{ if } \ \bar{n}_d = 0)$$

and the other two polynomials have a similar structure. Note that by the construction $\bar{A}^*(z)$ and $\bar{C}^*(z)$ do not have common factors. The reciprocal polynomials $\bar{A}(z)$ and $\bar{C}(z)$, defined in analogy with (2.3), will also be utilized.

Put

$$f(z) = \frac{\bar{A}C}{A\bar{C}}(z)\, \frac{C^*}{A^*\hat{C}^*\bar{C}^*}(z)\, z^k. \qquad (4.3)$$

It then follows from (4.1) that the conditions for the gradient to vanish can be written as

$$\frac{1}{2\pi i} \oint \bar{C}^*(z) f(z) z^j \frac{dz}{z} = 0, \qquad j = 1, \cdots, \hat{n}_a$$

$$\frac{1}{2\pi i} \oint \bar{A}^*(z) f(z) z^j \frac{dz}{z} = 0, \qquad j = 1, \cdots, \hat{n}_c. \qquad (4.4)$$

Furthermore introduce

$$F_j = \frac{1}{2\pi i} \oint f(z) z^j \frac{dz}{z} \qquad (4.5)$$

and (4.4) can be written as

$$\begin{bmatrix} 1 & \bar{c}_1 & \cdots & \bar{c}_{\bar{n}_c} & & & 0 \\ & \ddots & & & \ddots & & \\ 0 & & & 1 & \bar{c}_1 & \cdots & \bar{c}_{\bar{n}_c} \\ \hline 1 & \bar{a}_1 & \cdots & \bar{a}_{\bar{n}_a} & & & 0 \\ & \ddots & & & \ddots & & \\ 0 & & & 1 & \bar{a}_1 & \cdots & \bar{a}_{\bar{n}_a} \end{bmatrix} \begin{bmatrix} F_1 \\ \vdots \\ F_{\hat{n}_a + \hat{n}_c - \bar{n}_d} \end{bmatrix} = 0. \qquad (4.6)$$

The matrix is a $(\hat{n}_a + \hat{n}_c) \times (\hat{n}_a + \hat{n}_c - \bar{n}_d)$ matrix. Since $\bar{A}(z)$ and $\bar{C}(z)$ do not have common factors, it follows from the theory of resultants, see e.g., [6], that the matrix has rank $(\hat{n}_a + \hat{n}_c - \bar{n}_d)$. Thus, it is concluded

$$F_i = 0, \qquad i = 1, \cdots, (\hat{n}_a + \hat{n}_c - \bar{n}_d). \qquad (4.7)$$

To proceed the following result is needed.

Lemma 1: Consider the function

$$f(z) = \frac{g(z)}{\prod_{i=1}^{l} (z - u_i)^{t_i}} \qquad (4.8)$$

where g is analytic inside and on the unit circle, the numbers u_i are distinct and $t_i \geqslant 1$. Assume that

$$\frac{1}{2\pi i} \oint f(z) z^j \frac{dz}{z} = 0, \qquad j = 1, \cdots, m \qquad (4.9)$$

where the integration path is the unit circle and

$$m = \sum_{i=1}^{l} t_i. \qquad (4.10)$$

Then f is analytic inside the unit circle. □

A proof of the lemma is given in the Appendix.

Considering the form of the function f (4.3), where the integer k is defined by (2.10), it is clear that the cases $k > 0$ and $k \leqslant 0$ should be discussed separately. Observing that

$$z^k = z^{\max(0,k)} / z^{\max(0,-k)}$$

both cases can be treated simultaneously.

The function $f(z)$ is now rewritten as follows:

$$f(z) = \frac{\bar{A}CC^*}{A^*\hat{C}^*\bar{C}^*}(z) \cdot z^{\max(0,k)} \cdot \frac{1}{A\bar{C}(z) \cdot z^{\max(0,-k)}}. \qquad (4.11)$$

Note that the poles of $f(z)$ are precisely the zeros of $A\bar{C}(z) \cdot z^{\max(0,-k)}$. The number of them is

$$\begin{aligned} n_a + \bar{n}_c &+ \max(0, -n_a - \hat{n}_c + \hat{n}_a + n_c) \\ &= \max(n_a + \hat{n}_c, \hat{n}_a + n_c) - \bar{n}_d. \end{aligned} \qquad (4.12)$$

Furthermore identify

$$g(z) = \frac{\bar{A}CC^*}{A^*\hat{C}^*\bar{C}^*}(z)\cdot z^{\max(0,k)} \tag{4.13}$$

which clearly is analytic inside and on the unit circle.

In order to apply the lemma the condition (4.10) must be satisfied, which requires that

$$\hat{n}_a + \hat{n}_c - \bar{n}_d \geqslant \max(n_a + \hat{n}_c, \hat{n}_a + n_c) - \bar{n}_d$$

which can be rewritten as

$$\min(\hat{n}_a - n_a, \hat{n}_c - n_c) \geqslant 0 \tag{4.14}$$

which is nothing but (2.12).

Application of Lemma 1 thus implies that f is analytic inside the unit circle. Thus the zeros of $A\bar{C}(z)\cdot z^{\max(0,-k)}$ must be matched by zeros of the numerator. Since $C^*(z)$ has all its zeros outside the unit circle, it is found that $A\bar{C}(z)\cdot z^{\max(0,-k)}$ must be a factor of $\bar{A}C(z)z^{\max(0,k)}$. The degree of the last polynomial is

$$\bar{n}_a + n_c + \max(0,k) = \max(\hat{n}_a + n_c, n_a + \hat{n}_c) - \bar{n}_d \tag{4.15}$$

which is to be compared with (4.12). Thus, the two polynomials have the same degree. Moreover, the values of the leading coëfficients are both one. This implies that the polynomials are identical, i.e.,

$$A\bar{C}(z)\cdot z^{\max(0,-k)} = \bar{A}C(z)\cdot z^{\max(0,k)} \tag{4.16}$$

or rewritten with the reciprocal polynomials after multiplication with $\bar{D}(z)$

$$A^*(z)\hat{C}^*(z) = \hat{A}^*(z)C^*(z). \tag{4.17}$$

Applying Theorem 1 it is found that the local and global minimum points of the loss function coincide or more precisely:

Theorem 2: Consider the loss function $V: R^{\hat{n}_a} \times H(\hat{C}) \rightarrow R$ defined by (2.7). Let $\hat{n}_a \geqslant n_a$, $\hat{n}_c \geqslant n_c$. Then the gradient of V will vanish if and only if

$$A^*(z)\hat{C}^*(z) = \hat{A}^*(z)C^*(z). \qquad \square$$

Thus, it has been shown that the conditions (2.12) are sufficient for all local extrema to be global minima. The conditions $\hat{n}_a \geqslant n_a$ and $\hat{n}_c \geqslant n_c$ can not be removed as shown by the following example.

Example: Let the process be given by

$$y(t) + \alpha y(t-2) = e(t), \qquad \alpha \neq 0$$

and let $\hat{n}_a = 0$ and $\hat{n}_c = 1$. Clearly (2.12) is not fulfilled. V is a function of $\hat{c}$ only. It can be calculated to satisfy

$$2V(\hat{c}) = \frac{1 - \alpha\hat{c}^2}{(1-\alpha^2)[1 - (1-\alpha)\hat{c}^2 - \alpha\hat{c}^4]}. \tag{4.18}$$

It is easy to see that $V'(0) = 0$ $V''(0) = (1 - 2\alpha)/(1-\alpha^2)$, and thus, $\hat{c} = 0$ is a local *maximum* point if $\alpha > 0.5$. Since $\lim_{|\hat{c}|\rightarrow 1} V(\hat{c}) = \infty$, there exist (at least) two local minimum points.

Summing up the analysis, the following conclusions can be made. If $\hat{n}_a \geqslant n_a$ and $\hat{n}_c = n_c$, or $\hat{n}_a = n_a$ and $\hat{n}_c \geqslant n_c$, then there will be a unique global and local minimum. If $\hat{n}_a > n_a$ and $\hat{n}_c > n_c$, there will be many global and local minima. All minima are characterized by (3.11), which means that all models correspond to processes with the same spectral density. If $\hat{n}_a < n_a$ or $\hat{n}_c < n_c$ there may be several local minima.

Appendix
Proof of Lemma 1

Consider first the case when all t_i equals 1. Equation (4.9) can be written as

$$0 = \frac{1}{2\pi i}\oint \frac{g(z)}{\prod_{i=1}^{l}(z-u_i)} z^j \frac{dz}{z} = \sum_{k=1}^{l} \operatorname*{res}_{z=u_k} \frac{g(z)}{\prod_{i=1}^{l}(z-u_i)} \cdot z^{j-1} = \sum_{k=1}^{l} h_k u_k^{j-1}, \qquad j = 1,\cdots,m \tag{A.1}$$

where

$$h_k = \frac{g(u_k)}{\prod_{\substack{i=1\\ i\neq k}}^{l}(u_k - u_i)}. \tag{A.2}$$

Equations (A.1) can be written in matrix formulation as

$$\begin{bmatrix} 1 & \cdots & 1 \\ u_1 & & u_l \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ u_1^{m-1} & & u_l^{m-1} \end{bmatrix} \begin{bmatrix} h_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ h_l \end{bmatrix} = 0. \tag{A.3}$$

The matrix is a vanderMonde matrix. Since $\{u_i\}$ are distinct, its rank is l if $m \geqslant l$. Hence, $h_k = 0$ for $k = 1 \cdots l$, which from (A.2) implies that $g(u_k) = 0$ for $k = 1 \cdots l$. Thus, the poles of f in $u_1 \cdots u_l$ are cancelled and f is in fact analytic inside the unit circle.

In order to treat the general case proceed as follows. The formula (A.1) will be generalized to

$$\begin{aligned} 0 &= \frac{1}{2\pi i}\oint \frac{g(z)}{\prod_{i=1}^{l}(z-u_i)^{t_i}} z^{j-1}dz \\ &= \sum_{k=1}^{l} \operatorname*{res}_{z=u_k} \frac{g(z)}{\prod_{i=1}^{l}(z-u_i)^{t_i}} z^{j-1} \\ &= \sum_{k=1}^{l} \frac{1}{(t_k-1)!} D^{(t_k-1)}\left[\frac{g(z)}{\prod_{\substack{i=1\\ i\neq k}}^{l}(z-u_i)^{t_i}}\cdot z^{j-1}\right]_{z=u_k} \\ &= \sum_{k=1}^{l} \frac{1}{(t_k-1)!} \sum_{\nu=0}^{t_k-1} (t_{k_\nu} - 1) D^{(\nu)}[z^{j-1}]_{z=u_k} \\ &\quad \cdot D^{(t_k-1-\nu)}\left[\frac{g(z)}{\prod_{\substack{i=1\\ i\neq k}}^{l}(z-u_i)}\right]_{z=u_k} \qquad j = 1,\cdots,m. \end{aligned} \tag{A.4}$$

In matrix formulation these equations become

$$Ud = [U_1 \cdots U_l] \begin{bmatrix} d_1 \\ \cdot \\ \cdot \\ \cdot \\ d_l \end{bmatrix} = 0 \qquad (A.5)$$

where the block matrix U_k $(1 \leqslant k \leqslant l)$ is a $m \times t_k$ matrix given by

$$U_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ z & 1 & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ z^{m-1} & D[z^{m-1}] & \cdots & D^{(t_k-1)}[z^{m-1}] \end{bmatrix}_{z=u_k} \cdot \qquad (A.6)$$

The vector $d_k (1 \leqslant k \leqslant l)$ has t_k components and is given by

$$d_{k,j} = \frac{1}{(j-1)!(t_k - j)!} D^{(t_k - j)} \left[\frac{g(z)}{\prod_{\substack{i=1 \\ i \neq k}}^{l} (z - u_i)} \right]_{z=u_k}, \quad 1 \leqslant j \leqslant t_k. \qquad (A.7)$$

The matrix U is a generalization of the vanderMonde matrix. It follows from [5] that the assumption (4.10) implies that rank U equals $\sum_1^l t_i$. Thus, it is concluded from (A.5) that $d = 0$, or

$$D^{(t_k - j)} \left[\frac{g(z)}{\prod_{\substack{i=1 \\ i \neq k}}^{l} (z - u_i)^{t_i}} \right]_{z=u_k} = 0, \quad 1 \leqslant j \leqslant t_k,\ 1 \leqslant k \leqslant l. \qquad (A.8)$$

This relation can be written as

$$\sum_{\nu=0}^{t_k - j} (t_{k_\nu} - j) D^{(\nu)}[g(z)]_{z=u_k} \cdot D^{(t_k - j - \nu)} \left[\frac{1}{\prod_{\substack{i=1 \\ i \neq k}}^{l} (z - u_i)^{t_i}} \right]_{z=u_k} = 0, \quad 1 \leqslant j \leqslant t_k,\ 1 \leqslant k \leqslant l. \qquad (A.9)$$

Consider first (A.9) for $j = t_k$. Then $D^{(0)}[g(z)]_{z=u_k} = 0$ is easily derived. Consider then (A.9) for $j = t_k - 1$, which in a similar way implies $D^{(1)}[g(z)]_{z=u_k} = 0$. By repeating the discussion it can be concluded that

$$D^{\nu}[g(z)]_{z=u_k} = 0, \qquad 0 \leqslant \nu \leqslant t_k - 1,\ 1 \leqslant k \leqslant l. \qquad (A.10)$$

Equation (A.10) implies that the polynomial $\prod_{i=1}^{l}(z - u_i)^{t_i}$ divides the function $g(z)$. This means that all poles of the function f inside the unit circle are cancelled. With this observation the proof of the lemma is finished. □

References

[1] K. J. Åström, *Introduction to Stochastic Control Theory.* New York: Academic, 1970.

[2] K. J. Åström and T. Bohlin, "Numerical identification of linear dynamical systems from normal operating records," presented at the IFAC Symp. Theory of Self-Adaptive Systems, Teddington, England, 1965; also in *Theory of Self-Adaptive Control Systems*, P. H. Hammond, Ed. New York: Plenum, 1966.

[3] T. Bohlin, "Analysis of stationary EEG-signals by the maximum likelihood and generalized least-squares methods," IBM Nordic Laboratory, Sweden, Tech. Paper TP 18.200, 1971.

[4] E. J. Hannan, *Time Series Analysis.* London: Menthen, London, 1960.

[5] I. Kaufman, "The inversion of the vanderMonde matrix and the transformation to the Jordan canonical form," *IEEE Trans. Automat. Contr.* (Corresp.), vol. AC-14, pp. 774–777, Dec. 1969.

[6] B. L. van der Waerden, *Moderne Algebra I.* Berlin: Springer, 1937.

Karl J. Åström (M'71) was born in 1934. He received the Civ.Ing. degree in teknisk fysik in 1957 and the Tekn.Lic. in mathematics and automatic control in 1960, both from the Royal Institute of Technology (KTH), Stockholm, Sweden.

From 1955 to 1960 he held various teaching positions at the KTH. During 1957–1960 he worked on inertial guidance for the Research Institute of National Defense. The work resulted in a new mechanization of a navigation system which was successfully flight tested. In 1961 he joined the IBM Nordic Laboratory to work on theory and application of process control. He worked with optimal and stochastic control as a Visiting Scientist at IBM Control Systems Research, San José, Calif. from 1962 to 1963. From 1964 to 1965 he was manager of the Control Systems Theory Group at the IBM Nordic Laboratory with responsibilities to develop and implement strategies for on-line control of a paper machine. Since 1965 he has been Professor of the Chair of Automatic Control at the Lund Institute of Technology, Lund, Sweden. During the academic year 1969–1970 he was Visiting Professor of Applied Mathematics at Brown University. He has published two books *Reglerteori* (in Swedish) and *Introduction to Stochastic Control Theory* and various papers in the areas of optimal control, stochastic control systems, identification, adaptive control, and application of control theory.

Torsten Söderström was born in Malmö, Sweden in 1945. He received the M.S. degree (Civilingenjör) in applied physics and mathematics in 1969 and the Ph.D. degree in automatic control in 1973 from the Lund Institute of Technology, Lund, Sweden.

He is currently a Lecturer at the Lund Institute of Technology. His main research interests include various aspects of system identification.

Reprinted by permission from IEEE TRANSACTIONS ON AUTOMATIC CONTROL
Vol. AC-19, No. 6, December 1974, pp. 769-773