Nonlinear Maximum Likelihood Estimation of AR Time Series

L. Todd McWhorter L. L. Scharf
Department of Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309

Abstract

In this paper we describe an algorithm for finding the exact *maximum likelihood (ML) estimators for the parameters of an autoregressive time series. We demonstrate that the ML normal equations can be written as an interdependent set of cubic and quadratic equations in the AR polynomial coefficients. We present an algorithm, based on the properties of Sylvester resolvent matrices, that solves this set of non-linear equations for low-order problems.*

1 Introduction

In this paper we derive a new algorithm for computing maximum likelihood (ML) estimators of the parameters which characterize a stationary Gaussian autoregressive time series[1]. The derivation is based on the Gohberg-Semencul formula for the inverse of a Toeplitz matrix. Our key result is a set of equations we have labeled the normal equations of maximum likelihood, to distinguish them from the normal equations of linear prediction. The normal equations of maximum likelihood are at most cubic in the autoregressive parameters, whereas the normal equations of linear prediction are, of course, linear. We present an algebraically exact algorithm based on the properties of Sylvester resolvent matrices which is used to solve this set of non-linear equations.

Any attempt to summarize the vast literature on autoregressive modeling, or identification of autoregressive time series, would be futile. In the remainder of this section, we briefly describe the context of the results in this paper. A more detailed summary of the existing literature on this subject can be found in [9].

With reference to Table 1, we may organize work on ML autoregressive (AR) modeling according to the criterion for identification and the representation used to describe the AR model. These are the columns and rows of Table 1.

Table 1: Classification of Algorithms by Criterion and Representation of $\mathbf{R}^{-1}$

Repr. for R, R^{-1}, or $A(z)$.	Criterion	
	RML	ML
Toeplitz only	NA	Burg, et al. [4]
Levinson	Vis-Scharf [12]	Morf, et al. [10] Dugre, et al. [5]
Gohberg-semencul	Kay [8]	Kailath, et al. [7] Box-Jenkins [3] McWhorter-Scharf eq. (20)
Markovian	NA	Schweppe [11] Akaike [1] Ansley [2]

In column 1, the work of Kay [8] is difficult to classify because it uses two representations for the AR time series. That is, it uses a Gohberg-Semencul characterization of $\mathbf{R}^{-1}$, but it uses a Levinson formula to represent the AR model in the recursive maximization scheme. We classify this work as recursive maximum likelihood (RML), with a Gohberg-Semencul formula for the inverse correlation matrix. The work of Vis and Scharf [12] is classified as RML with a Levinson formula for the representation of $\mathbf{R}^{-1}$ and the order increasing AR models.

In column 2, the theory of exact maximum likelihood (ML) estimation of AR parameters begins with the work of Schweppe [11], although he provided no algorithms for the maximization of likelihood. Akaike [1] and Ansley [2] did provide such algorithms. This work is classified as ML, based on a Markovian representation for the time series and its correlation sequence.

The work of Morf, et al. [10] provided a link between Markovian representations and Levinson recursions, leading to the formulas of Dugre, et al. [5] for computing likelihood. Neither [10] nor [5] contained

[1] This work was supported by Bonneville Power Administraton under contract #DEBI7990BPO7346 and by the Office of Naval Research, Statistics and Probability Branch, under contract #N00014-89-J1070.

formulas for maximizing likelihood.

The work of Kailath, et al. [7], Box and Jenkins [3], and this paper are classified as ML, based on the Gohberg-Semencul formula for the inverse correlation matrix. No algorithms were presented in [7] or [3] for maximizing likelihood. Our contribution is to maximize likelihood by deriving a new set of non-linear normal equations and to present algorithms for solving them. The work of Burg, et al. [4] is not exact ML because it use ML to estimate a Toeplitz correlation matrix without assuming a model for the time series. Approximate ML estimates of the AR coefficients are then obtained by solving the normal equations of linear prediction, using the estimated correlation matrix.

In summary, we offer a new method for maximizing exact likelihood for AR time series, based on the Gohberg-Semencul formulas for the inverse of a Toeplitz matrix. We introduce a set of non-linear normal equations, which illustrate how the ML estimate of the AR model constructs estimated correlation matrices that are much different than those of linear prediction. We also present an algorithm for solving the normal equations which is applicable for low-order problems.

2 Normal Equations

Let $\mathbf{y} = [y_0 \;\; y_1 \;\; \cdots \;\; y_{N-1}]^{\mathrm{T}}$ be a snapshot of an AR time series. That is, assume the time series is synthesized by passing a real-valued, zero-mean white Gaussian noise process, of variance σ^2, through the AR filter $1/A(z)$. The p^{th} order polynomial $A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$ is monic and minimum phase. The coefficients $\{a_i\}_1^p$ are real-valued. These assumptions imply that $\mathbf{y}$ is distributed $N[\mathbf{0}, \mathbf{R}]$ where $\mathbf{R} \in \mathbb{R}^{N \times N}$ is the symmetric Toeplitz correlation matrix associated with the AR time series. We denote the northwest $(p+1) \times (p+1)$ block of $\mathbf{R}$ by $\mathbf{R}_p$.

Assume that we are given M statistically independent snapshots $\mathbf{Y} = [\mathbf{y}_1 \; \mathbf{y}_2 \; \cdots \; \mathbf{y}_M]$ of a time series synthesized as just described. We assume that each snapshot has N elements. The log-likelihood function for the data can be written as

$$L = -\frac{MN}{2}\ln(2\pi) - \frac{M}{2}\ln(|\mathbf{R}|) - \frac{M}{2}\mathrm{tr}\{\mathbf{R}^{-1}\mathbf{S}\}$$

where

$$\mathbf{S} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{y}_i\mathbf{y}_i^{\mathrm{T}}$$

is the sample correlation matrix. The correlation matrix, $\mathbf{R}$, can be described in terms of the AR coefficients and the input noise variance. The Gohberg-Semencul inversion formulae, described in [6], provide one such link between the correlation matrix and the AR coefficients. The key Gohberg-Semencul formula in our derivation is

$$\mathbf{R}^{-1} = \frac{1}{\sigma^2}\mathbf{Q}^{-1} = \frac{1}{\sigma^2}(\mathbf{F}\mathbf{F}^{\mathrm{T}} - \mathbf{G}\mathbf{G}^{\mathrm{T}}) \qquad (1)$$

where $\mathbf{F}$ and $\mathbf{G}$ are the $N \times N$ lower triangular Toeplitz matrices

$$\mathbf{F} = \sum_{i=0}^{p} a_i \mathbf{Z}_i; \quad \mathbf{G} = \sum_{i=1}^{p} a_i \mathbf{Z}_{N-i}; \quad (a_0 = 1).$$

The $N \times N$ shift matrices $\mathbf{Z}_k$ are zero except for ones on their k^{th} sub-diagonal. That is,

$$\mathbf{Z}_k = [\delta(k, i-j)]_{i,j}; \quad \mathbf{Z}_0 = \mathbf{I}; \quad \mathbf{Z}_N = \mathbf{0}.$$

Observe that the data dependent term of the log-likelihood function can be written as

$$\begin{aligned}\mathrm{tr}\{\mathbf{R}^{-1}\mathbf{S}\} &= \frac{1}{\sigma^2}\mathrm{tr}\{[\mathbf{F}\mathbf{F}^{\mathrm{T}} - \mathbf{G}\mathbf{G}^{\mathrm{T}}]\mathbf{S}\} \\ &= \frac{1}{\sigma^2}\mathrm{tr}\left[\sum_{i=0}^{p} a_i \mathbf{Z}_i^{\mathrm{T}}\right]\mathbf{S}\left[\sum_{j=0}^{p} a_j \mathbf{Z}_j\right] \\ &- \frac{1}{\sigma^2}\mathrm{tr}\left[\sum_{i=0}^{p} a_i \mathbf{Z}_{N-i}^{\mathrm{T}}\right]\mathbf{S}\left[\sum_{j=0}^{p} a_j \mathbf{Z}_{N-j}\right]. \qquad (2)\end{aligned}$$

Let $\mathbf{a} = [1 \; a_1 \; \cdots \; a_p]^{\mathrm{T}}$ be the vector of AR coefficients and define the $(p+1) \times (p+1)$ matrix $\hat{\mathbf{R}}_p(\mathbf{Y}) = [\hat{r}_{ij}(\mathbf{Y})]$ where

$$\hat{r}_{ij}(\mathbf{Y}) = \frac{1}{N}\left[\mathrm{tr}\{\mathbf{Z}_i^{\mathrm{T}}\mathbf{S}\mathbf{Z}_j\} - \mathrm{tr}\{\mathbf{Z}_{N-i}^{\mathrm{T}}\mathbf{S}\mathbf{Z}_{N-j}\}\right]. \qquad (3)$$

Then equation (2), in conjunction with equation (3), implies

$$\mathrm{tr}\{\mathbf{R}^{-1}\mathbf{S}\} = \frac{N}{\sigma^2}\mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a}. \qquad (4)$$

The log-likelihood function, ignoring irrelevant constants, can be now be written as

$$L = -\frac{MN}{2}\ln(\sigma^2) - \frac{M}{2}\ln(|\mathbf{Q}|) - \frac{MN}{2\sigma^2}\mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a}.$$

Equation (4) has been derived by Kay [8] and is fundamental in the development of the RML algorithm described in his paper. Kay has shown that $\hat{\mathbf{R}}_p(\mathbf{Y})$ is not, in general, non-negative definite. In [9], it is also shown that

$$E\{\hat{\mathbf{R}}_p(\mathbf{Y})\} = \mathbf{R}_p - \mathbf{B} \qquad (5)$$

where

$$\mathbf{B} = \frac{1}{N}\begin{bmatrix} 0 & r_1 & \cdots & pr_p \\ r_1 & 2r_0 & \cdots & (p+1)r_{p-1} \\ 2r_2 & 3r_1 & \cdots & (p+2)r_{p-2} \\ \vdots & \vdots & & \vdots \\ pr_p & (p+1)r_{p-1} & \cdots & 2pr_0 \end{bmatrix}. \quad (6)$$

This equation indicates that $\hat{\mathbf{R}}_p(\mathbf{Y})$ is a biased estimator of $\mathbf{R}_p$, with bias $\mathbf{B}$.

The normal equations of maximum likelihood can be derived by differentiating the Lagrangian

$$\begin{aligned}\mathcal{L} &= -\frac{MN}{2}\ln(\sigma^2) - \frac{M}{2}\ln(|\mathbf{Q}|) - \frac{MN}{2\sigma^2}\mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} \\ &\quad -\lambda(\mathbf{a}^{\mathrm{T}}\boldsymbol{\delta} - 1) \quad (7)\end{aligned}$$

with respect to the unknown parameters. The vector $\boldsymbol{\delta} = [1\ 0\ \cdots\ 0]^{\mathrm{T}}$ is the first column of the $(p+1) \times (p+1)$ identity matrix. Differentiating equation (7) with respect to σ^2 yields the normal equation

$$-\frac{MN}{2\sigma^2} + \frac{MN}{2\sigma^4}\mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} = 0.$$

Solving this equation for σ^2 produces the maximum likelihood estimate for σ^2:

$$\hat{\sigma}^2 = \mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a}. \quad (8)$$

If the estimate of $A(z)$ is minimum phase, then the maximum likelihood estimate of σ^2 will be non-negative despite the fact that $\hat{\mathbf{R}}_p(\mathbf{Y})$ is not always non-negative definite [8].

To find the ML estimator of the AR coefficients, first observe that

$$\frac{\partial}{\partial a_i}\ln(|\mathbf{Q}|) = \mathrm{tr}\{\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial a_i}\} = -\mathrm{tr}\{\mathbf{Q}\frac{\partial \mathbf{Q}^{-1}}{\partial a_i}\},$$

and

$$\begin{aligned}\frac{\partial \mathbf{Q}^{-1}}{\partial a_i} &= \frac{\partial}{\partial a_i}(\mathbf{F}\mathbf{F}^{\mathrm{T}} - \mathbf{G}\mathbf{G}^{\mathrm{T}}) \\ &= \mathbf{Z}_i\mathbf{F}^{\mathrm{T}} + \mathbf{F}\mathbf{Z}_i^{\mathrm{T}} - \mathbf{Z}_{N-i}\mathbf{G}^{\mathrm{T}} - \mathbf{G}\mathbf{Z}_{N-i}^{\mathrm{T}}.\end{aligned}$$

After some algebra, it can be shown that

$$\begin{aligned}&\frac{\partial}{\partial a_i}\ln(|\mathbf{Q}|) = \\ &= -\frac{2}{\sigma^2}\sum_{j=0}^{p} a_j \mathrm{tr}\{\mathbf{Z}_j^{\mathrm{T}}\mathbf{R}\mathbf{Z}_i - \mathbf{Z}_{N-j}^{\mathrm{T}}\mathbf{R}\mathbf{Z}_{N-i}\}. \quad (9)\end{aligned}$$

Define the row vector $\mathbf{c}_i^{\mathrm{T}} = [c_{i0}\ c_{i1}\ \cdots\ c_{ip}]$ and the matrix $\mathbf{C}$ where

$$c_{ij} = \mathrm{tr}\{\mathbf{Z}_j^{\mathrm{T}}\mathbf{R}\mathbf{Z}_i\} - \mathrm{tr}\{\mathbf{Z}_{N-j}^{\mathrm{T}}\mathbf{R}\mathbf{Z}_{N-i}\} \quad (10)$$

and

$$\mathbf{C}^{\mathrm{T}} = [\mathbf{c}_0\ \mathbf{c}_1\ \cdots\ \mathbf{c}_p]. \quad (11)$$

Equation (9) and the definitions of equations (10) and (11) can be used to obtain

$$\frac{\partial}{\partial a_i}\ln(|\mathbf{Q}|) = -\frac{2}{\sigma^2}\mathbf{c}_i^{\mathrm{T}}\mathbf{a}; \quad \frac{\partial}{\partial \mathbf{a}}\ln(|\mathbf{Q}|) = -\frac{2}{\sigma^2}\mathbf{C}\mathbf{a}.$$

We are now in a position to derive the normal equations for the ML estimators of the AR coefficients. The gradient of the Lagrangian in equation (7) with respect to $\mathbf{a}$, when equated to zero, yields the normal equations

$$\begin{aligned}&\tfrac{M}{\sigma^2}\mathbf{C}\mathbf{a} - \tfrac{MN}{\sigma^2}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} = \lambda\boldsymbol{\delta} \\ &[\tfrac{1}{N}\mathbf{C} - \hat{\mathbf{R}}_p(\mathbf{Y})]\mathbf{a} = \tfrac{\lambda\sigma^2}{MN}\boldsymbol{\delta}. \quad (12)\end{aligned}$$

From equations (3), (5), and (10), it can be seen that the $(p+1) \times (p+1)$ matrix $\mathbf{C}$ is equivalent to

$$\mathbf{C} = N(\mathbf{R}_p - \mathbf{B})$$

where $\mathbf{B}$ is defined in equation (6). The ML normal equations can now be written as

$$[\mathbf{R}_p - \mathbf{B} - \hat{\mathbf{R}}_p(\mathbf{Y})]\mathbf{a} = \frac{\lambda\sigma^2}{MN}\boldsymbol{\delta}. \quad (13)$$

Recall that for an AR time series, $\mathbf{R}_p\mathbf{a} = \sigma^2\boldsymbol{\delta}$. Therefore equation (13) can be reduced to

$$[\hat{\mathbf{R}}_p(\mathbf{Y}) + \mathbf{B}]\mathbf{a} = \sigma^2(1 - \frac{\lambda}{MN})\boldsymbol{\delta}. \quad (14)$$

The important non-linear term in the normal equations is the vector $\mathbf{B}\mathbf{a}$. To simplify this expression, note that the bias matrix of equation (6), can be written as

$$\mathbf{B} = \frac{1}{N}(\mathbf{D}\mathbf{R}_p + \mathbf{R}_p\mathbf{D})$$

where $\mathbf{D} = \mathrm{diag}\{0, 1, 2, \cdots, p\}$. Again, the relation $\mathbf{R}_p\mathbf{a} = \sigma^2\boldsymbol{\delta}$ can be invoked to write

$$\begin{aligned}\mathbf{B}\mathbf{a} &= \frac{1}{N}(\mathbf{D}\mathbf{R}_p + \mathbf{R}_p\mathbf{D})\mathbf{a} = \frac{1}{N}\mathbf{D}\boldsymbol{\delta}\sigma^2 + \frac{1}{N}\mathbf{R}_p\mathbf{D}\mathbf{a} \\ &= \frac{1}{N}\mathbf{R}_p\mathbf{a}_d\end{aligned}$$

where $\mathbf{a}_d = \mathbf{D}\mathbf{a} = [0\ a_1\ 2a_2\ \cdots\ pa_p]^{\mathrm{T}}$. The normal equations can now be reduced to

$$\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} + \frac{1}{N}\mathbf{R}_p\mathbf{a}_d = \sigma^2(1 - \frac{\lambda}{MN})\boldsymbol{\delta}. \quad (15)$$

The Lagrange multiplier can now be easily determined from equation (15):

$$\begin{aligned}&\mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} + \tfrac{1}{N}\mathbf{a}^{\mathrm{T}}\mathbf{R}_p\mathbf{a}_d = \sigma^2(1 - \tfrac{\lambda}{MN}) \\ &\sigma^2 + \tfrac{\sigma^2}{N}\boldsymbol{\delta}^{\mathrm{T}}\mathbf{a}_d = \sigma^2(1 - \tfrac{\lambda}{MN}) \quad \Rightarrow \lambda = 0.\end{aligned}$$

The ML normal equations can then be reduced to

$$[\hat{\mathbf{R}}_p(\mathbf{Y}) + \mathbf{B}]\mathbf{a} = \sigma^2\boldsymbol{\delta} \tag{16}$$

or

$$\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} + \frac{1}{N}\mathbf{R}_p\mathbf{a}_d = \sigma^2\boldsymbol{\delta}. \tag{17}$$

Equation (16) provides an intuitive connection between the theory of exact maximum likelihood and the least squares theory of linear prediction. In the least squares theory, the normal equations are $\hat{\mathbf{R}}_{LP}\mathbf{a} = \sigma^2\boldsymbol{\delta}$ where $\hat{\mathbf{R}}_{LP}$ is an estimate of the correlation matrix $\mathbf{R}_p$. Thus linear prediction and exact maximum likelihood share a common structure in their respective normal equations. Linear prediction builds a "reasonable" estimate of the correlation matrix solely from the data and then finds the optimal whitening polynomial. Whereas exact maximum likelihood builds a deficient estimate of the correlation matrix and *simultaneously* tries to offset the deficiencies and whiten the "corrected" estimate of the correlation matrix.

In the remainder of this section we manipulate the ML normal equations into a form that makes them amenable to either exact or iterative solution. From equation (17) we obtain

$$\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} + \frac{1}{N}\sigma^2\mathbf{Q}_p\mathbf{a}_d = \sigma^2\boldsymbol{\delta}$$

or

$$\mathbf{Q}_p^{-1}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} + \frac{1}{N}\sigma^2\mathbf{a}_d = \sigma^2\mathbf{Q}_p^{-1}\boldsymbol{\delta} = \sigma^2\mathbf{a}. \tag{18}$$

The matrix $\mathbf{Q}_p$ is the $(p+1)\times(p+1)$ northwest block of the matrix $\mathbf{Q}$ defined in the Gohberg-Semencul formula of equation (1). This matrix is within a scale constant of $\mathbf{R}_p$. Recall that our maximum likelihood estimate of σ^2 is $\mathbf{a}^{\mathrm{T}}\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a}$. Therefore equation (18) can be written as

$$[\mathbf{Q}_p^{-1} - \mathbf{a}\mathbf{a}^{\mathrm{T}} + \frac{1}{N}\mathbf{a}_d\mathbf{a}^{\mathrm{T}}]\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} = \mathbf{0}. \tag{19}$$

If we define the $(p+1)\times(p+1)$ analogs of $\mathbf{F}$ and $\mathbf{G}$, we arrive at

$$[\mathbf{F}_p\mathbf{F}_p^{\mathrm{T}} - \mathbf{G}_p\mathbf{G}_p^{\mathrm{T}} - \mathbf{a}\mathbf{a}^{\mathrm{T}} + \frac{1}{N}\mathbf{a}_d\mathbf{a}^{\mathrm{T}}]\hat{\mathbf{R}}_p(\mathbf{Y})\mathbf{a} = \mathbf{0}. \tag{20}$$

The first element of this vector equation is satisfied for all $\mathbf{a}$ and $\hat{\mathbf{R}}_p(\mathbf{Y})$. We are left with p equations that must be solved for the p AR coefficients $\{a_i\}_1^p$. We will show in the next section that the normal equations can be written as an interdependent set of cubic and quadratic equations in the AR coefficients. We also present algorithms to solve this set of equations.

3 Exact ML Algorithm

In this section we present an algorithm for solving the normal equations of equation (20). We call this section *exact* solution methods because we characterize the solutions for the AR coefficients as roots of polynomials. In practice there will be errors associated with any root finding algorithm, so the results will be exact only in an algebraic sense. A discussion of iterative algorithms which solve the normal equations can be found in [9]. In the following we present the exact algorithm for solving the normal equations of a third-order system. The algorithm for a general p^{th}-order system is merely an extension of the procedure given below.

To begin, use the normal equations of (20) to generate the three polynomials

$$p_1(x)|_{x=a_2} = a_2^3\alpha_3 + a_2^2\alpha_2 + a_2\alpha_1 + \alpha_0 = 0 \tag{21}$$

$$p_2(x)|_{x=a_2} = a_2^2\beta_2 + a_2\beta_1 + \beta_0 = 0 \tag{22}$$

$$p_3(x)|_{x=a_2} = a_2^2\gamma_2 + a_2\gamma_1 + \gamma_0 = 0 \tag{23}$$

where the coefficients $\{\alpha_i\}$, $\{\beta_i\}$, and $\{\gamma_i\}$ are functions of the AR coefficients a_1 and a_3 and the data through the matrix $\hat{\mathbf{R}}_p(\mathbf{Y})$. To find the maximum likelihood solution we must solve this interdependent set of equations. One can interpret the normal equations (21), (22), and (23) as three polynomials which share a root at $x = a_2$. It is this interpretation which indicates that the properties of the Sylvester resolvent matrix might be applicable in the solution of the normal equations. We first describe the construction of the Sylvester resolvent matrix before solving this system of equations.

Let $a(x)$ and $b(x)$ be the polynomials

$$a(x) = \sum_{i=0}^{n} a_i x^i = a_n \prod_{i=1}^{n}(x - \alpha_i) \quad a_n \neq 0$$

$$b(x) = \sum_{i=0}^{m} b_i x^i = b_m \prod_{i=1}^{m}(x - \beta_i) \quad b_m \neq 0$$

and define the vector $\boldsymbol{\psi}_k(x) = [1 \;\; x \;\; x^2 \;\; \cdots \;\; x^{k-1}]^{\mathrm{T}}$. The Sylvester resolvent matrix, $\mathbf{M}$, can be implicitly defined by

$$\mathbf{M}\boldsymbol{\psi}_{n+m}(x) = \begin{bmatrix} a(x)\boldsymbol{\psi}_m(x) \\ b(x)\boldsymbol{\psi}_n(x) \end{bmatrix}.$$

The properties of the resolvent matrix are described in more detail in [9]. It can be shown that the determinant of the Sylvester resolvent matrix is

$$\det(\mathbf{M}) = b_m^n \prod_{k=1}^{m} a(\beta_k).$$

Thus, $\det(\mathbf{M}) = 0$ if and only if the polynomials $a(x)$ and $b(x)$ share at least one root. It is this property which we will exploit in the algorithm that follows.

Form a Sylvester resolvent matrix,

$$\mathbf{M}_1 = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & 0 \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_0 & \beta_1 & \beta_2 & 0 & 0 \\ 0 & \beta_0 & \beta_1 & \beta_2 & 0 \\ 0 & 0 & \beta_0 & \beta_1 & \beta_2 \end{bmatrix},$$

for the polynomials $p_1(x)$ and $p_2(x)$. Similarly, form a Sylvester resolvent matrix, $\mathbf{M}_2$, for the polynomials $p_1(x)$ and $p_3(x)$. The normal equations imply that the polynomials $p_1(x)$ and $p_2(x)$ share a common root at a_2. The same assertion is true for $p_1(x)$ and $p_3(x)$. The properties of the Sylvester resolvent matrix can then be used to obtain the two polynomials

$$\begin{aligned} p_4(a_1) &= \det(\mathbf{M}_1) = a_1^q \lambda_q + \cdots + a_1\lambda_1 + \lambda_0 = 0 \\ p_5(a_1) &= \det(\mathbf{M}_2) = a_1^r \xi_r + \cdots + a_1\xi_1 + \xi_0 = 0. \end{aligned}$$

The coefficients $\{\lambda_i\}$ and $\{\xi_i\}$ are now only functions of a_3 and the data $\hat{\mathbf{R}}_p(\mathbf{Y})$. These two equations indicate that the polynomials $p_4(x)$ and $p_5(x)$ share a root at a_1. Therefore the Sylvester resolvent matrix, $\mathbf{M}_3$, for the polynomials $p_4(x)$ and $p_5(x)$ must be singular. This fact can be used to determine our last equation

$$p_6(a_3) = \det(\mathbf{M}_3) = a_3^s \kappa_s + \cdots + a_3\kappa_1 + \kappa_0 = 0$$

where now the coefficients of this polynomial depend exclusively on the data $\hat{\mathbf{R}}_p(\mathbf{Y})$. The AR coefficient a_3 must be a root of the polynomial $p_6(x)$. The procedure is then to find all roots of $p_6(x)$ that are real and have magnitude less than one. All roots that satisfy this property are potential solutions for a_3. Potential a_1 solutions are the real roots of either $p_4(x)$ or $p_5(x)$ with the coefficients of these polynomials formed from the data and the potential solutions for a_3. The solutions for a_2 can be found in a similar fashion from the polynomials $p_1(x)$, $p_2(x)$, or $p_3(x)$. This procedure generates a finite number of potential solution sets for the AR coefficients. The final step of the procedure is to eliminate all sets of solutions that do not generate minimum phase $A(z)$ and/or do not satisfy all three normal equations.

It is straightforward to extend this procedure for systems of arbitrary order. However, because the polynomials generated by this technique increase in order exponentially, this procedure is only tractable for low-order systems. In [9] we present an iterative algorithm that can be applied to higher-order systems.

References

[1] Akaike, H.,"Maximum likelihood identification of Gaussian autoregressive moving average models," *Biometrika*, vol. 60, pp 255-265, 1973.

[2] Ansley, G. F., "An algorithm for the exact likelihood of a mixed autoregressive moving average process," *Biometrika*, vol. 66, pp 59-65, 1979.

[3] Box, G. E. P., and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

[4] Burg, J. P., D. G. Luenburger, and D. L. Wenger, "Estimation of structured covariance matrices," *IEEE Proc.*, vol. 70, pp 963-974, Sept. 1982.

[5] Dugre, J. P., L. L. Scharf, and C. Gueguen, "Exact likelihood for stationary autoregressive moving average processes," *Signal Processing*, vol. 11, pp 105-118, 1986.

[6] Heinig, Georg, and Karla Rost, *Algebraic methods for Toeplitz-like matrices and operators*," Birkhauser, 1984.

[7] Kailath, T., B. Levy, L. Ljung, and M. Morf, "Fast time invariant implementation of Gaussian signal detectors," *IEEE Trans. IT*, vol. 24, no. 4, July 1978.

[8] Kay, Steven M., "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. ASSP*, vol. ASSP-31:1, pp. 56-65, Feb. 1983.

[9] McWhorter, L. Todd, and L. L. Scharf, "Exact maximum likelihood identification of autoregressive time series," *IEEE Trans. SP*, (accepted for publication).

[10] Morf, M., G. S. Sidhu, and T. Kailath, "Some new algorithms for recursive estimation on constant, linear discrete time systems," *IEEE Trans. AC*, vol. 19, pp 315-323, Aug. 1974.

[11] Schweppe, F. C., "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. IT*, vol. 11, pp 61-70, 1965.

[12] Vis, M. L., and L. L. Scharf, "A note on recursive maximum likelihood for autoregressive modeling," *IEEE Trans. SP*, to be published 1994.