

## A Note on Recursive Maximum Likelihood for Autoregressive Modeling

Marvin L. Vis and Louis L. Scharf

**Abstract**—In this paper, we rederive recursive maximum likelihood (RML) for an autoregressive (AR) time series using the Levinson decomposition. This decomposition produces a recursive update of the likelihood function for the AR parameters in terms of the reflection coefficients, prediction error variances, and forward and backward prediction errors. A fast algorithm for this recursive update is presented and compared with the recursive updates of the Burg algorithm. The comparison clarifies the connection between Burg's algorithm and RML.

### I. INTRODUCTION

Begin with the multivariate normal density for the normal random variable  $\mathbf{x}$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{R}_{xx}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{R}_{xx}^{-1} \mathbf{x} \right\}. \quad (1)$$

The vector  $\mathbf{x}$  is an  $N$ -length snapshot of a  $p$ th-order autoregressive, AR( $p$ ), time series

$$\{x_t\} = \frac{\sigma^2}{A(z)} \{u_t\} \quad (2)$$

$$A(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p}, \quad (3)$$

where  $\{u_t\}$  is a white Gaussian noise (WGN) sequence with zero mean and unit variance. We want to find the set of AR parameters  $\{a_i\}_1^p$  that maximizes the probability of the occurrence of  $\mathbf{x}$ .

The approach to AR parameter estimation advocated by Kay [1] is to "recursively maximize" exact likelihood by assuming that the  $n$ th iteration ( $n \leq p$ ) that the time series  $\mathbf{x}$  comes from an AR( $n$ ) process and that the set of reflection coefficients  $\{k_i\}_1^{n-1}$  is known. Then, the density  $f_n(\mathbf{x})$  can be manipulated into a function of data  $\{x_i\}_0^{n-1}$ , known parameters  $\{k_i\}_1^{n-1}$ , and  $k_n$ . Thus parameterized,  $f_n(\mathbf{x})$  may be maximized with respect to  $k_n$ . In this correspondence, we have derived a fast algorithm for recursively updating the order-increasing likelihood function by using the Levinson decomposition of  $\mathbf{R}_{xx}^{-1}$ .

### II. NEW DERIVATION OF RML

The strategy we use to obtain the likelihood function is to first assume that the snapshot  $\mathbf{x} = [x_0 x_1 \dots x_{N-1}]^T$  is the output of an AR( $n$ ) filter driven by unit variance WGN. Next, using the Levinson recursions and their inverse LDU factorization of the covariance matrix  $\mathbf{R}_{xx}$ , we write the density  $f_n(\mathbf{x})$  in terms of  $\mathbf{x}$ , the  $n$ th-order AR parameters and the prediction error variance  $\sigma_n^2$ . From this, we get the likelihood function, which we compress with respect to  $\sigma_n^2$  to get the desired compressed likelihood function  $\mathcal{L}_n(\mathbf{a}_n)$ , where  $\mathbf{a}_n^T = [a_n^n \dots a_1^n 1]$ . We can then express the AR( $n$ ) parameters of  $\mathbf{a}_n$  in terms of the AR( $n-1$ ) parameters of  $\mathbf{a}_{n-1}$  and the  $n$ th reflection coefficient  $k_n$  using the Levinson recursions

$$\mathbf{a}_n = (\mathbf{I} + k_n \mathbf{J}) \begin{bmatrix} 0 \\ \mathbf{a}_{n-1} \end{bmatrix} \quad (4)$$

Manuscript received October 4, 1993; revised February 10, 1994. This work was supported by Bonneville Power Administration under Contract no. DEB17990BP07346 and by the Office of Naval Research, Statistics, and Probability Branch under Contract no. N00014-89-J-1070.

The authors are with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425.

IEEE Log Number 9403749.

where  $\mathbf{I}$  is the identity matrix, and  $\mathbf{J}$  is the exchange matrix

$$\mathbf{J} = \begin{bmatrix} \mathbf{0} & & & 1 \\ & \ddots & & \\ & & \ddots & \\ 1 & & & \mathbf{0} \end{bmatrix}. \quad (5)$$

We will also use the Levinson recursion for the prediction error variances:

$$\sigma_n^2 = (1 - k_n^2) \sigma_{n-1}^2. \quad (6)$$

Assuming that  $\mathbf{x}$  comes from an AR( $n$ ) process, it can be shown that the  $i$ th reflection coefficient  $k_i$  is zero for  $i > n$ , and therefore, the LDU factorization of  $\mathbf{R}_{f_n} = \frac{1}{\sigma_n^2} \mathbf{R}_{xx}$  is

$$\begin{aligned} \mathcal{A}_n^T \mathbf{R}_{f_n} \mathcal{A}_n &= \mathbf{D}_n^2 \\ \mathbf{R}_{f_n}^{-1} &= \mathcal{A}_n \mathbf{D}_n^{-2} \mathcal{A}_n^T \\ \mathbf{D}_n^2 &= \frac{1}{\sigma_n^2} \text{diag}(\sigma_0^2, \dots, \sigma_{n-1}^2, \sigma_n^2, \dots, \sigma_n^2) \\ &= \begin{bmatrix} \mathbf{D}_n^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ \mathcal{A}_n^T &= \left[ \begin{array}{cccc|cccc} 1 & & & & & & & \\ a_1^1 & 1 & \mathbf{0} & & & & & \mathbf{0} \\ \vdots & & \ddots & & & & & \\ a_{n-1}^{n-1} & \dots & a_1^{n-1} & 1 & & & & \\ \hline a_n^n & \dots & \dots & a_1^n & 1 & & & \\ \mathbf{0} & \ddots & & & \ddots & \ddots & & \mathbf{0} \\ & & & a_n^n & \dots & \dots & a_1^n & 1 \end{array} \right] \\ &= \begin{bmatrix} \hat{\mathbf{A}}_{n-1}^T & \mathbf{0} \\ \mathbf{A}_n^T \end{bmatrix}. \end{aligned} \quad (7)$$

Using this decomposition, we can write the density for  $\mathbf{x}$ , which is now indexed by the assumed order  $n$ , as

$$f_n(\mathbf{x}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}} |\mathbf{D}_n^2|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma_n^2} \mathbf{x}^T \mathcal{A}_n \mathbf{D}_n^{-2} \mathcal{A}_n^T \mathbf{x} \right\}. \quad (8)$$

Taking log-likelihood, compressing with respect to  $\sigma_n^2$ , and ignoring constants, we get the compressed likelihood formula

$$\mathcal{L}_n = -\ln |\mathbf{D}_n^2| - N \ln [Q_n(\mathbf{x})] \quad (9)$$

where  $Q_n(\mathbf{x})$  is the quadratic form

$$Q_n(\mathbf{x}) = \mathbf{x}^T \mathcal{A}_n \mathbf{D}_n^{-2} \mathcal{A}_n^T \mathbf{x}. \quad (10)$$

The idea behind RML is to manipulate  $\mathcal{L}_n$  into a function of the "known" parameters  $\{k_i\}_1^{n-1}$  and  $\mathbf{x}$  and the "unknown" parameter  $k_n$  so that we can maximize  $\mathcal{L}_n$  with respect to  $k_n$ . The recursion for the determinant term  $-\ln |\mathbf{D}_n^2|$  is

$$-\ln |\mathbf{D}_n^2| = n \ln(1 - k_n^2) - \ln |\mathbf{D}_{n-1}^2|, \quad \ln |\mathbf{D}_0| = 0. \quad (11)$$

Now, consider the quadratic form  $Q_n(\mathbf{x}) = \|\mathbf{D}_n^{-1} \mathcal{A}_n^T \mathbf{x}\|^2$ . The original derivation of RML in [1] used a formula found in [2] to write out this quadratic as  $\mathbf{a}_n^T \mathbf{S}_n \mathbf{a}_n$ , but here, we instead use the Levinson decomposition of  $\mathbf{R}_{f_n}$  to obtain an equivalent expression for the quadratic. Because we have fixed  $\{k_i\}_1^{n-1}$  (and thus have also fixed  $\{a_i^j\}_{j=1}^n$  for  $j = 1, 2, \dots, n-1$ ),  $\hat{\mathbf{A}}_{n-1}$  is known, and

we can write

$$\mathbf{A}_n^T \mathbf{x} = \begin{bmatrix} \hat{\mathbf{A}}_{n-1}^T & \mathbf{0} \\ \mathbf{A}_n^T & \mathbf{0} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \hat{\varepsilon}_n \\ \mathbf{X}_n \mathbf{a}_n \end{bmatrix}, \quad (12)$$

where  $\hat{\varepsilon}_n$  is known, and  $\mathbf{X}_n$  is the Hankel data matrix for the covariance method of linear prediction assuming an  $n$ th-order predictor:

$$\mathbf{X}_n = \begin{bmatrix} x_0 & x_1 & \cdots & x_n \\ x_1 & \cdots & x_n & x_{n+1} \\ \vdots & \ddots & \vdots & \vdots \\ x_n & & & \vdots \\ \vdots & & & x_{N-2} \\ x_{N-n-1} & \cdots & x_{N-2} & x_{N-1} \end{bmatrix}. \quad (13)$$

(Note:  $\mathbf{X}_n \in R^{(N-n) \times (n+1)}$  so that its dimension changes with order  $n$ .) Furthermore

$$\begin{aligned} \mathbf{D}_n^{-1} \mathbf{A}_n^T \mathbf{x} &= \begin{bmatrix} \hat{\mathbf{D}}_n^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_n \\ \mathbf{X}_n \mathbf{a}_n \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{D}}_n^{-1} \hat{\varepsilon}_n \\ \mathbf{X}_n \mathbf{a}_n \end{bmatrix}, \\ \hat{\varepsilon}_n &= \begin{bmatrix} \hat{\varepsilon}_{n-1} \\ \alpha_{n-1} \end{bmatrix} \\ \alpha_{n-1} &= [x_0 \quad x_1 \quad \cdots \quad x_{n-1}] \mathbf{a}_{n-1}. \end{aligned} \quad (14)$$

This means we may write

$$\|\hat{\mathbf{D}}_n^{-1} \hat{\varepsilon}_n\|^2 = (1 - k_n^2) \left( \|\hat{\mathbf{D}}_{n-1}^{-1} \hat{\varepsilon}_{n-1}\|^2 + \alpha_{n-1}^2 \right). \quad (15)$$

Now, let's write the quadratic  $\mathbf{a}_n^T \mathbf{X}_n^T \mathbf{X}_n \mathbf{a}_n$  in terms of the  $n$ th reflection coefficient  $k_n$  and the known parameters in  $\mathbf{a}_{n-1}$  using the Levinson recursion of (4). Substituting for  $\mathbf{a}_n$ , we have

$$\|\mathbf{X}_n \mathbf{a}_n\|^2 = \left\| \mathbf{X}_n (\mathbf{I} + k_n \mathbf{J}) \begin{bmatrix} 0 \\ \mathbf{a}_{n-1} \end{bmatrix} \right\|^2. \quad (16)$$

Equation (16) is simply a quadratic function of  $k_n$  and the forward and backward prediction errors of the order  $n-1$  predictor. Define the  $n$ th-order forward and backward prediction errors as

$$\begin{aligned} \mathbf{e}_n &= \begin{bmatrix} \alpha_n \\ \hat{\mathbf{e}}_n \end{bmatrix} = \mathbf{X}_n \mathbf{a}_n \\ &= \mathbf{X}_n (\mathbf{I} + k_n \mathbf{J}) \begin{bmatrix} 0 \\ \mathbf{a}_{n-1} \end{bmatrix} \\ &= \hat{\mathbf{e}}_{n-1} + k_n \hat{\mathbf{f}}_{n-1} \\ \mathbf{f}_n &= \begin{bmatrix} \hat{\mathbf{f}}_n \\ \beta_n \end{bmatrix} = \mathbf{X}_n \mathbf{J} \mathbf{a}_n \\ &= \mathbf{X}_n \mathbf{J} (\mathbf{I} + k_n \mathbf{J}) \begin{bmatrix} 0 \\ \mathbf{a}_{n-1} \end{bmatrix} \\ &= \hat{\mathbf{f}}_{n-1} + k_n \hat{\mathbf{e}}_{n-1}. \end{aligned} \quad (17)$$

Now, we have the following expression for  $\mathbf{a}_n^T \mathbf{X}_n^T \mathbf{X}_n \mathbf{a}_n$ :

$$\mathbf{a}_n^T \mathbf{X}_n^T \mathbf{X}_n \mathbf{a}_n = \|\mathbf{e}_n\|^2 = \|\hat{\mathbf{e}}_{n-1} + k_n \hat{\mathbf{f}}_{n-1}\|^2. \quad (18)$$

Thus, the quadratic form in (10) can be written as

$$\begin{aligned} Q_n(\mathbf{x}) &= \|\hat{\mathbf{D}}_n^{-1} \hat{\varepsilon}_n\|^2 + \|\mathbf{e}_n\|^2 \\ &= (1 - k_n^2) \left( \|\hat{\mathbf{D}}_{n-1}^{-1} \hat{\varepsilon}_{n-1}\|^2 + \alpha_{n-1}^2 \right) \\ &\quad + \|\hat{\mathbf{e}}_{n-1} + k_n \hat{\mathbf{f}}_{n-1}\|^2. \end{aligned} \quad (19)$$

However, since  $\mathbf{R}_{xx}$  and  $\mathbf{R}_{xx}^{-1}$  are  $\mathbf{J}$  symmetric,  $\mathbf{x}^T \mathbf{R}_{xx}^{-1} \mathbf{x} =$

$\hat{\mathbf{x}}^T \mathbf{R}_{xx}^{-1} \hat{\mathbf{x}}$ , where

$$\hat{\mathbf{x}} = \mathbf{J} \mathbf{x} = [x_{N-1} \quad x_{N-2} \quad \cdots \quad x_0]^T. \quad (20)$$

Therefore, we can also write the quadratic form  $Q_n(\mathbf{x})$  as

$$\begin{aligned} Q_n(\mathbf{x}) &= \|\hat{\mathbf{D}}_n^{-1} \hat{\varepsilon}_n\|^2 + \|\mathbf{f}_n\|^2 \\ &= (1 - k_n^2) \left( \|\hat{\mathbf{D}}_{n-1}^{-1} \hat{\varepsilon}_{n-1}\|^2 + \beta_{n-1}^2 \right) \\ &\quad + \|\hat{\mathbf{f}}_{n-1} + k_n \hat{\mathbf{e}}_{n-1}\|^2. \end{aligned} \quad (21)$$

$$\hat{\varepsilon}_n = \hat{\mathbf{A}}_{n-1}^T \begin{bmatrix} x_{N-1} \\ x_{N-2} \\ \vdots \\ x_{N-n-1} \end{bmatrix}.$$

Equations (19) and (21) are simply two different ways to write the quadratic form  $Q_n(\mathbf{x})$ . Therefore, by summing them and dividing the sum by two, we get the following:

$$\begin{aligned} Q_n(\mathbf{x}) &= \frac{1}{2} \left\{ \|\hat{\mathbf{D}}_n^{-1} \hat{\varepsilon}_n\|^2 + \|\hat{\mathbf{D}}_{n-1}^{-1} \hat{\varepsilon}_{n-1}\|^2 + V_n \right\}, \\ V_n &= \|\mathbf{e}_n\|^2 + \|\mathbf{f}_n\|^2. \end{aligned} \quad (22)$$

The first two terms of (22), which we will call  $h_n^2$ , can be written using the recursive step

$$\begin{aligned} h_n^2 &= \|\hat{\mathbf{D}}_n^{-1} \hat{\varepsilon}_n\|^2 + \|\hat{\mathbf{D}}_{n-1}^{-1} \hat{\varepsilon}_{n-1}\|^2 \\ &= (1 - k_n^2) (h_{n-1}^2 + \alpha_{n-1}^2 + \beta_{n-1}^2). \end{aligned} \quad (23)$$

The remainder of (22) ( $V_n$ ) is the sum of squared forward and backward prediction errors. It is this sum that Burg [3] minimized in his estimator. Following his derivation, write this sum as

$$\begin{aligned} V_n(k_n) &= \|\mathbf{e}_n\|^2 + \|\mathbf{f}_n\|^2 \\ &= (1 + k_n^2) \left( \|\hat{\mathbf{e}}_{n-1}\|^2 + \|\hat{\mathbf{f}}_{n-1}\|^2 \right) \\ &\quad + 4k_n \left( \hat{\mathbf{e}}_{n-1}^T \hat{\mathbf{f}}_{n-1} \right) \\ &= (1 + k_n^2) d_{n-1} + 4k_n c_{n-1}. \end{aligned} \quad (24)$$

As pointed out by Burg [3], we can put a recursive step on  $d_n$  to avoid calculating the sum of squared errors:

$$d_n = V_n(k_n) - \alpha_{n-1}^2 - \beta_{n-1}^2. \quad (25)$$

Putting together the results of (23) and (24), we get  $Q_n(\mathbf{x})$  as a quadratic in  $k_n$ :

$$Q_n(\mathbf{x}) = \frac{1}{2} [(1 - k_n^2) (h_{n-1}^2 + \alpha_{n-1}^2 + \beta_{n-1}^2) + V_n(k_n)]. \quad (26)$$

Finally, substituting (11) and (26) back into the compressed likelihood formula of (9) and ignoring terms that are constants with respect to  $k_n$ , we have an update formula for  $\mathcal{L}_n$  as a function of  $k_n$ :

$$\begin{aligned} \mathcal{L}_n(k_n) &= n \ln(1 - k_n^2) \\ &\quad - N \ln \left[ (1 - k_n^2) (h_{n-1}^2 + \alpha_{n-1}^2 + \beta_{n-1}^2) \right. \\ &\quad \left. + V_n(k_n) \right]. \end{aligned} \quad (27)$$

This is our key finding. It shows that Burg's sum of forward and backward prediction error variances  $V_n(k_n) = \|\mathbf{e}_n\|^2 + \|\mathbf{f}_n\|^2$  accounts for only part of the quadratic dependence of likelihood on the data  $\mathbf{x}$  and accounts for none of the determinant dependence on  $k_n$ .

The derivative  $\mathcal{L}'_n(k_n)$  with respect to  $k_n$  is cubic as reported in [1]. Furthermore, it can be shown, using the Gohberg-Semencul

formulas [4] for the inverse of a Toeplitz matrix, that the above recursive update on compressed likelihood is equivalent to the recursive update derived in [1]. However, if we consider the calculations required for finding the RMLE of the  $p$ th-order AR parameters from an  $N$ -dimensional snapshot, we find that the algorithm in [1] requires  $\mathcal{O}(p^2N)$  operations in addition to rooting  $p$  third-order polynomials, whereas our new approach requires  $\mathcal{O}(pN)$  operations in addition to rooting the  $p$  polynomials. The dominant  $p^2N$  operations for the algorithm in [1] are a result of having to compute the coefficients  $S_{ij}$  for the expression  $\mathbf{x}^T \mathbf{R}_{f_n}^{-1} \mathbf{x} = \mathbf{a}^T \mathbf{S} \mathbf{a}$ , which we avoid by using instead recursions on the forward and backward prediction errors.

III. RML ALGORITHM

Summarizing the recursive updates for the compressed likelihood formula, we have the following RML algorithm:

*RML Algorithm:*

**To Compute:**

RMLE of  $\mathbf{a}_p, \sigma_p^2$ , given  $p, \mathbf{x}$

**Initialization:**

$$\begin{aligned} \mathbf{a}_0 &= 1 \\ \sigma_0^2 &= \frac{1}{N} \mathbf{x}^T \mathbf{x} \\ \mathbf{e}_0 &= \begin{bmatrix} \alpha_0 \\ \hat{\mathbf{e}}_0 \end{bmatrix} = \mathbf{x} \\ \mathbf{f}_0 &= \begin{bmatrix} \hat{\mathbf{f}}_0 \\ \beta_0 \end{bmatrix} = \mathbf{x} \\ c_0 &= \hat{\mathbf{e}}_0^T \hat{\mathbf{f}}_0 \\ d_0 &= 2N\sigma_0^2 - \alpha_0^2 - \beta_0^2 \\ h_0^2 &= 0 \\ k_0 &= 0 \end{aligned}$$

**Body:**

For  $n = 1$  to  $p$ , do

$$\begin{aligned} V_n(k_n) &= (1 + k_n^2)d_{n-1} + 4k_n c_{n-1} \\ \mathcal{L}_n(k_n) &= n \ln(1 - k_n^2) - N \ln[(1 - k_n^2) \\ &\quad \times (h_{n-1}^2 + \alpha_{n-1}^2 + \beta_{n-1}^2) + V_n(k_n)] \end{aligned}$$

Root  $\mathcal{L}'_n(k_n) \rightarrow k_n$

$$\mathbf{a}_n = (\mathbf{I} + k_n \mathbf{J}) \begin{bmatrix} 0 \\ \mathbf{a}_{n-1} \end{bmatrix}$$

$$\sigma_n^2 = (1 - k_n^2)\sigma_{n-1}^2$$

$$d_n = V_n(k_n) - \alpha_{n-1}^2 - \beta_{n-1}^2$$

$$h_n^2 = (1 - k_n^2)(h_{n-1}^2 + \alpha_{n-1}^2 + \beta_{n-1}^2)$$

$$\begin{bmatrix} \mathbf{e}_n \\ - \\ \mathbf{f}_n \end{bmatrix} = \begin{bmatrix} \alpha_n \\ \hat{\mathbf{e}}_n \\ \beta_n \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I} & k_n \mathbf{J} \\ k_n \mathbf{I} & \mathbf{J} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{e}}_{n-1} \\ - \\ \hat{\mathbf{f}}_{n-1} \end{bmatrix}$$

$$c_n = \hat{\mathbf{e}}_n^T \hat{\mathbf{f}}_n$$

(end loop on  $n$ )

IV. COMPARISON TO BURG'S ALGORITHM

Our update formula obtains the same answer as Kay's update, but it does it differently, using a formula that is similar to Burg's. In fact, Burg's update for forward and backward prediction error variances

is embedded in our recursion. The philosophy for recursive update common to RML and Burg is to recursively estimate the  $n$ th reflection coefficient  $k_n$  by assuming that the set of previously estimated reflection coefficients  $\{k_i\}_1^{n-1}$  (or, equivalently, the AR coefficients  $\{a_i^j\}_{i=1}^j$  for  $j = 1, 2, \dots, n-1$ ) are known. The RML approach is to choose  $k_n$  such that the compressed likelihood function is maximized under this assumption. The Burg approach is to minimize

$$\begin{aligned} V_n(k_n) &= \|\mathbf{e}_n\|^2 + \|\mathbf{f}_n\|^2 \\ &= \|\hat{\mathbf{e}}_{n-1} + k_n \hat{\mathbf{f}}_{n-1}\|^2 + \|k_n \hat{\mathbf{e}}_{n-1} + \hat{\mathbf{f}}_{n-1}\|^2 \end{aligned} \tag{28}$$

with respect to  $k_n$ .

As we saw in the derivation of RML,  $V_n(k_n)$  appears as a portion of the quadratic form in the likelihood function. Hence, Burg's algorithm updates only a part of the compressed likelihood function. It does not account for the determinant term nor does it completely account for the quadratic form, whereas RML updates the compressed likelihood function.

V. CONCLUSION

We have derived a fast algorithm for recursive maximum likelihood (RML) of AR parameters using the Levinson decomposition of the correlation matrix  $\mathbf{R}_{x,x}$ . The Levinson decomposition is a natural approach for AR maximum likelihood in the sense that the decomposition is given in terms of the prediction errors and the order-increasing AR whiteners. Because the  $i$ th reflection coefficient is zero for  $i > n$  in this decomposition (assuming an AR( $n$ ) process), the  $n$ th recursive update can be calculated using the  $(n-1)$ st forward and backward prediction errors. Thus, the order-increasing AR whiteners are never used explicitly as they are in Kay's algorithm for RML.

This derivation, together with the findings in [1], provide a means to recursively fit higher order AR models to data using a fast algorithm. Further, as pointed out in [1], the estimated AR model is guaranteed to be stable. When we compare RML to the Burg estimator, we find that the philosophy used by both estimators for recursively updating the order-increasing models is consistent, but the Burg algorithm optimizes only a portion of the compressed likelihood function.

REFERENCES

- [1] S. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-31, no. 1, Feb. 1983.
- [2] G. E. P. Box and G. M. Jenkins, *Time Series Analysis-Forecasting and Control*. San Francisco, CA: Holden Day, 1970.
- [3] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. Dissertation, Stanford Univ., Stanford, CA, 1975.
- [4] T. Kailath, A. Vieira, and M. Morf, "Inverses of Toeplitz operators, innovations, and orthogonal polynomials," *SIAM Rev.*, vol 20, no. 1, Jan. 1978.