AN ADAPTIVE ARMA SPECTRAL ESTIMATOR: PART 1*

James A. Cadzow and Randolph L. Moses
Department of Electrical and Computer Engineering
Arizona State University
Tempe, Arizona 85281

Abstract - In this two part paper, a novel procedure for generating an ARMA spectral model of a wide sense stationary time series is developed. The parameters of this model are selected so that they most closely fit a set of Yule-Walker equations which are estimated from a finite set of time series' observations. This ARMA modeling method has been found to exhibit a spectral estimation performance which is typically superior to such alternatives as the maximum entropy (AR) method, classical Fourier procedures (MA), and, the Box-Jenkins method (ARMA).

One of the principal features of this spectral estimation method is the elegant algebraic structure of the linear system of equations which need be solved when finding the ARMA model's parameters. This shift-invariant type structure gives rise to an adaptive algorithmic solution procedure whose computational efficiency is comparable to that achieved by recently developed fast AR algorithmic methods. The details of the adaptive ARMA modeling procedure will be covered in Part 2 of this paper. These dual characteristics of excellent estimation performance and real time adaptive implementation mark this method as being a primary spectral estimation tool.

I. INTRODUCTION

In many interdisciplinary applications, it is desired to estimate the essential attributes of a generally complex valued wide-sense stationary time series $\{x(n)\}$. Depending on the specific nature of the time series, this characterization is often adequately revealed through knowledge of the time series' associated autocorrelation sequence

$$r_x(n) = E\{x(n+m)x^*(m)\} \qquad n=0, \pm 1, \pm 2, \ldots \tag{1}$$

in which E and * denote the operations of expectation and complex conjugation, respectively. On the other hand, the requisite characterization may often be better made in the frequency domain through the spectral density function

$$S_x(e^{j\omega}) = \sum_{n=-\infty}^{\infty} r_x(n)e^{-j\omega n} \tag{2}$$

which is recognized as being the Fourier transform of the autocorrelation sequence. Either member of this transform pair conveys the total second-order statistical information relative to the underlying time series. Frequently, this second order statistical characterization provides all the information required for a given application (e.g., optimal Wiener filtering, one-step prediction, etc.).

The classical spectral estimation problem is concerned with estimating the spectral density function (2) from a finite set of time series observations. Without loss of generality, these observations will be taken to be the following N contiguous elements

$$x(1), x(2), \ldots, x(N) \tag{3}$$

A variety of procedures have been proposed for using these observations to effect a spectral density estimate. Invariably, the resultant estimate will take on a rational model form as expressed by

$$\hat{S}_x(e^{j\omega}) = \left| \frac{b_0 + b_1 e^{-j\omega} + \ldots + b_q e^{-jq\omega}}{1 + a_1 e^{-j\omega} + \ldots + a_p e^{-jp\omega}} \right|^2 = \left| \frac{B_q(e^{j\omega})}{A_p(e^{j\omega})} \right|^2 \tag{4}$$

in which the a_k and b_k are referred to as the model's autoregressive and moving average coefficients, respectively. We shall refer to this particular rational form as an autoregressive-moving average (ARMA) model of order (p,q). It is well known that any continuous spectral density can be approximated arbitrarily closely by this rational form if the order pair (p,q) is selected adequately large. Thus, by imposing a rational form on the spectral model, we incur no real loss in spectral representation.

The preponderance of research and application interest has been focused on two special cases of the above ARMA model. They are the moving average (MA) model in which all of the a_k coefficients are set to zero, and, the autoregressive (AR) model for which all of the b_k coefficients except b_0 are set to zero. The spectral density estimate arising from a MA model is seen to possess no poles, and as such it is frequently referred to as an all-zero model. Using similar reasoning, the AR model is referred to as an all-pole model, and, the general ARMA model is referred to as a pole-zero model.

Classical Fourier approaches [1] and the periodogram method [2] are procedures which ultimately provide a MA spectral density model. Similarly, the maximum entropy method and linear predictive coding are techniques that result in AR spectral density models. Undoubtedly, the primary reasons for interest in special case MA and AR models lie in the fact that they: (i) are amenable to a tractable analysis, (ii) typically provide adequate spectral estimation performance, and (iii) give rise to coefficient selection procedures which are implementable by computationally efficient algorithms.

Despite this predisposition towards MA and AR models, a growing interest in ARMA models is evident [3]-[9]. This is in recognition of the fact that the more general ARMA model usually provides superior spectral estimation performance while at the same time requires fewer model parameters to achieve that behavior. It is because of these very factors that a number of ARMA modeling procedures have been proposed. These include the Box-Jenkins maximum likelihood method [3], whitening filter approaches [4], [5], and, more recently, Cadzow's high performance method [6]-[9]. This latter method has been found to provide a spectral estimation performance which typically excels that obtained from its MA, AR, and ARMA counterparts.

* This work was supported in part by the Office of Naval Research, Statistics and Probability Program under Contract N00014-80-C-0303.

In this paper, we first characterize the modeling of a pure ARMA time series. An analytical procedure is presented for determining the underlying a_k and b_k coefficients in which the time series' actual autocorrelation element values are used. This idealistic situation then provides the justification for introducing the high performance method in which the ARMA model's coefficients are estimated from time series observations and not from autocorrelation values. It is shown that the p autoregressive a_k coefficients are obtained by solving a consistent system of p linear equations. When using this direct approach, the complete set of time series observations (3) are incorporated to effect a single spectral estimate in one computational effort. This approach is typically referred to as "block processing". Moreover, by using the generalized Levinson algorithm [10]-[11], it is possible to solve the above mentioned system of linear equations in a computationally efficient manner.

In Part 2 of this paper, a recursive procedure is developed in which the ARMA model's coefficients are updated as each new time series observation becomes available. In this "time-update processing" mode, an adaptive form of spectral estimation is thereby achieved. One of the particularly attractive features of this time-updating mode is its computational efficiency. Specifically, the p autoregressive coefficients (in actuality prediction errors) are optimally updated with each new time series observation. The number of multiplication and addition computations required in this updating is of the order p. Thus, the computational complexity of the high performance ARMA method is competitive with recently developed "fast" AR methods, but, its spectral estimation performance is typically far superior. The time-update mode is particularly attractive in those situations in which the time series being characterized is a long ongoing process and one wishes to generate a time evolving sequence of spectral estimates in a real time setting.

II. ARMA TIME SERIES: PERFECT MODELING

In this section, the second-order statistical characterization of an ARMA time series will be presented. This characterization will play a central role in the high performance spectral estimation procedure that is to be developed in subsequent sections. The time series $\{x(n)\}$ is said to be an ARMA time series of order (p,q) if it is generated according to the causal linear recursive relationship

$$x(n) = \sum_{k=0}^{q} b_k w(n-k) - \sum_{k=1}^{p} a_k x(n-k) \qquad (5)$$

in which $\{w(n)\}$ is a zero mean white noise excitation whose individual elements have variance one. It is readily shown that the spectral density corresponding to the response time series $\{x(n)\}$ is given by expression (4). Thus, there is seen to be an equivalence between a rational spectral density model and the response of a causal linear system to a white noise excitation.

We will now direct our attention to developing a systematic procedure for identifying the recursive system's autoregressive coefficients (i.e., the a_k) and moving average coefficients (i.e., the b_k) from the response time series' autocorrelation elements. It will be beneficial to consider separately the tasks of identifying these two different sets of coefficients.

Autoregressive Coefficient Identification

The autoregressive coefficients can be determined directly upon examining the autocorrelation characterization of recursive system (5). This is achieved by first multiplying both sides of this recursive expression by $x^*(n-m)$ and then taking the expected value. This is found to result in the well known Yule-Walker equations

$$\sum_{k=1}^{p} a_k r_x(m-k) = -r_x(m) \qquad \text{for } m \geq q+1 \qquad (6)$$

where it is important to note that the lag parameter m is here restricted to exceed the numerator order parameter q. As a side note, the Yule-Walker equations will involve the moving average coefficients b_k in a nonlinear manner for lags $0 \leq m \leq q$. The characteristic equations of expression (6) provide a straightforward procedure for obtaining the ARMA model's a_k autoregressive coefficients. This formally entails expressing the first "t" Yule-Walker equations (i.e., $q+1 \leq m \leq q+t$) in the following matrix format

$$\begin{bmatrix} r_x(q) & r_x(q-1) & \cdots & r_x(q-p+1) \\ r_x(q+1) & r_x(q) & \cdots & r_x(q-p+2) \\ \vdots & \vdots & & \vdots \\ r_x(q+t-1) & r_x(q+t-2) & \cdots & r_x(q-p+t) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \\ a_p \end{bmatrix} = - \begin{bmatrix} r_x(q+1) \\ r_x(q+2) \\ \vdots \\ r_x(q+t) \end{bmatrix} \qquad (7)$$

in which the integer t is taken to be equal to or larger than the model's denominator order (i.e., $t \geq p$). This linear system of equations may be compactly expressed as

$$R^q_{tp}\, \underline{a}_p = -\underline{r}^q_t \qquad (8)$$

where R^q_{tp} is a t×p autocorrelation matrix, $\underline{r}^q_t$ is a t×1 autocorrelation vector, and, $\underline{a}_p$ is the ARMA model's p×1 autoregressive coefficient vector. In this representation the subscripts t and p are appended to designate the number of Yule-Walker equations being used, and, the ARMA model's denominator order, respectively. Similarly, the superscript q depicts the ARMA model's numerator order.

To obtain the ARMA model's autoregressive coefficients, one then simply solves the consistent system of linear equations (8). Valuable insight relative to rational spectral density modeling is provided upon closer examination of the autocorrelation matrix's (i.e., R^q_{tp}) algebraic structure. It is convenient to express this characterization in the following theorem.

> **Theorem 1:** Let $\{r_x(k)\}$ designate the autocorrelation sequence which is associated with an ARMA time series of order (p,q). The corresponding system of t linear equations in m unknowns as specified by
>
> $$R^n_{tm}\, \underline{a}_m = \underline{r}^n_t \qquad (9)$$
>
> has a unique solution provided that m=p and n>q for any value of t≥p. Moreover, the rank of the t×m matrix R^n_{tm} is given by min (m,p,t) provided that n≥q and, by min (m,t) for 0≤n<q.

A proof of this theorem will not be given here, since these results are implicitly documented in various textbooks and papers dealing with time series. It is important to note that even if one has perfect autocorrelation knowledge of an ARMA time series, the evaluation of the associated autoregressive coefficients entails a determination of the order pair (p,q). This ordering information is implicitly contained in the algebraic structure of the autocorrelation matrix R^n_{tm}, and, can be obtained by examining this

structure for various combinations of the nonnegative integers m and n.

Moving Average Coefficient Determination

To determine the b_k coefficients associated with the ARMA time series, it will be beneficial to introduce the causal image of the time series' autocorrelation sequence as defined by

$$r_x^+(n) = r_x(n)u(n) - \frac{1}{2}r_x(0)\delta(n) \tag{10}$$

in which u(n) and $\delta(n)$ denote the standard unit-step and unit-Kronecker delta sequences, respectively. The autocorrelation sequence may be recovered from its causal image by using the complex conjugate symmetry property of autocorrelation sequences (i.e., $r_x(-n) = r_x^*(n)$). This reconstruction rule takes the form

$$r_x(n) = r_x^+(n) + r_x^+(-n)^* \tag{11}$$

Upon taking the Fourier transform of relationship (11), we have the required spectral density expression

$$S_x(e^{j\omega}) = S_x^+(e^{j\omega}) + S_x^+(e^{j\omega})^*$$

$$= 2\mathrm{Re}[S_x^+(e^{j\omega})] \tag{12}$$

where $S_x^+(e^{j\omega})$ denotes the Fourier transform of the causal image sequence $\{r_x^+(n)\}$.

In what is to follow, a parametric procedure for representing $S_x^+(ej\omega)$ (and therefore $S_x(ej\omega)$) will be given. This will first necessitate the introduction of the auxiliary sequence

$$c(n) = r_x^+(n) + \sum_{k=1}^{p} a_k r_x^+(n-k) \quad , \quad 0 \le n \le \max(q,p) \tag{13}$$

in which the causal autocorrelation elements as generated by relationship (10) and the autoregressive coefficients as obtained upon solving the system of equations (8) are used. According to the Yule-Walker equations (6) and the causal image definition (10), it is seen that this auxiliary sequence is identically zero outside the indexing range $0 \le n \le \max(q,p)$. With this in mind, the Fourier transform of relationship (13) is next taken and results in

$$C_s(e^{j\omega}) = \sum_{n=0}^{s} c(n)e^{-j\omega n} \tag{14a}$$

$$= [1 + \sum_{n=1}^{p} a_n e^{-j\omega n}]S_x^+(e^{j\omega})$$

$$= A_p(e^{j\omega})S_x^+(e^{j\omega}) \tag{14b}$$

in which $s = \max(q,p)$. Upon solving this relationship for $S_x^+(e^{j\omega})$ and substituting this solution into expression (12), the desired ARMA spectral density is obtained

$$S_x(e^{j\omega}) = \frac{C_s(e^{j\omega})}{A_p(e^{j\omega})} + \frac{C_s^*(e^{j\omega})}{A_p^*(e^{j\omega})}$$

$$= \frac{A_p^*(e^{j\omega})C_s(e^{j\omega}) + A_p(e^{j\omega})C_s^*(e^{j\omega})}{A_p(e^{j\omega})\,A_p^*(e^{j\omega})} \tag{15}$$

In order to determine the ARMA model's b_k moving average coefficients, we next use this relationship in conjunction with expression (4) to obtain

$$B_q(e^{j\omega})B_q^*(e^{j\omega}) = A_p(e^{j\omega})C_s^*(e^{j\omega}) + A_p^*(e^{j\omega})C_s(e^{j\omega}) \tag{16}$$

A spectral factorization of this expression will then yield the prerequisite b_k coefficients (assuming a minimum phase $B_q(e^{j\omega})$).

In summary, the spectral density and the associated a_k and b_k coefficients which characterize the ARMA time series of order (p,q) may be determined by following the four step procedure as outlined in Table 1. To carry out this model identification scheme, it is seen that knowledge of the order pair (p,q) and the q+p+1 autocorrelation elements $r_x(0)$, $r_x(1),\ldots,r_x(q+p)$ need be available.

1. Solve relationship (8) for the p autoregressive a_k coefficients. This will require setting $t \ge p$.
2. Generate the auxiliary sequence c(n) and its Fourier transform using expressions (13) and (14a), respectively.
3. The desired spectral density is then given by expression (15).
4. Perform a spectral factorization of the polynomial $B_q(e^{j\omega})B_q^*(e^{j\omega})$ as given by equation (16) to obtain the minimum phase choice of the b_k coefficients.

Table 1: Generation of the spectral density and the ARMA model parameters associated with a given set of autocorrelation values.

III. HIGH PERFORMANCE METHOD OF ARMA SPECTRAL MODELING

It is possible to adapt many of the ideas of Section II to achieve an ARMA spectral estimate when only the time series observations (3) are available (and not autocorrelation values). We shall again treat separately the cases of autoregressive and moving average coefficient determination.

Autoregressive Coefficient Estimation

To implement the autoregressive coefficient selection process as represented by relationship (8) it will be necessary to compute appropriate autocorrelation estimates from the given set of time series' observations. The high performance ARMA method effects these estimates in the guise of a convenient matrix format which lends itself to a particularly efficient computational realization [6]-[9]. In particular, the autocorrelation matrix and vector required in expression (8) are estimated according to

$$\hat{R}_{tp}^{q} = Y^{\dagger}X \tag{17}$$

$$\underline{\hat{r}}_t^{\,q} = Y^{\dagger}\underline{x} \tag{18}$$

where the dagger symbol † denotes the operation of complex conjugate transposition. The (N-p)×p Toeplitz type matrix X is specified by

$$X = \begin{bmatrix} x(p) & x(p-1) & \cdots & x(1) \\ x(p+1) & x(p) & \cdots & x(2) \\ \vdots & \vdots & & \vdots \\ x(N-1) & x(N-2) & & x(N-p) \end{bmatrix} \tag{19}$$

while the (N-p)×t Toeplitz type matrix Y has the form

$$Y = \begin{bmatrix} x(p-q) & x(p-q-1) & \cdots & x(p-q-t+1) \\ x(p-q+1) & x(p-q) & \cdots & x(p-q-t+2) \\ \vdots & \vdots & & \vdots \\ x(N-q-1) & x(N-q-2) & \cdots & x(N-q-t) \end{bmatrix} \quad (20)$$

and $\underline{x}$ is an $(N-p)\times 1$ vector given by[1]

$$\underline{x} = [x(p+1), x(p+2), \ldots x(N)]' \quad (21)$$

In formulating matrix Y, we have used the convention of setting to zero any elements x(k) for which k lies outside the observation index range $1 \le k \le N$.

If the autocorrelation matrix and vector estimates (17) and (18), respectively, are substituted into the Yule-Walker relationship (8), however, it is generally found that the resultant system of t equations in the p autoregressive coefficients is inconsistent for $t>p$. This is due to inevitable inaccuracies in the autocorrelation estimates, and, to a possible improper ARMA model order choice. In any case, the system of equations with these estimate substitutions will give rise to the $t\times 1$ Yule-Walker approximation error vector as specified by

$$\underline{e} = Y^{\dagger} X \underline{a} + Y^{\dagger} \underline{x} \quad (22)$$

Upon taking the expected value of $\underline{e}$, it is found that for the ARMA modeling order choice in which $q \ge p$, that this expectation results in

$$E\{e(k)\} = (N-q-k)\left[r_x(q+k) + \sum_{m=1}^{p} a_m r_x(q+k-m)\right], \quad 1 \le k \le t \quad (23)$$

while for the modeling order case $q<p$ this expectation produces

$$E\{e(k)\} = \begin{cases} (N-p)\left[r_x(q+k) + \sum_{m=1}^{p} a_m r_x(q+k-m)\right], & 1 \le k \le p-q \\ (N-q-k)\left[r_x(q+k) + \sum_{m=1}^{p} a_m r_x(q+k-m)\right], & p-q < k \le t \end{cases} \quad (24)$$

In either ordering case, it is seen that when the time series is an ARMA process of order (p,q), the expected value of the error vector $\underline{e}$ can be made equal to zero by a proper choice of the autoregressive coefficient vector $\underline{a}$. Namely, this selection would be such that the underlying Yule-Walker equations (6) are satisfied.[2] This implies that the system of equations (22) with $\underline{e} = \underline{\theta}$ provides an unbiased and a consistent estimate of the Yule-Walker equations (8), where $\underline{\theta}$ is the zero vector.

With the above thoughts in mind, an appealing approach to selecting the autoregressive coefficient vector is immediately suggested. Namely, $\underline{a}$ is chosen so as to make the error vector "as close" to its expected value of $\underline{\theta}$ as possible. This is of course predicated on the assumption that the time series is an ARMA process of order (p,q) or less. In order to attain a tractable procedure for selecting an appropriate autoregressive coefficient vector, we shall introduce the following quadratic functional

$$f(\underline{a}) = \underline{e}^{\dagger} \Lambda \underline{e} \quad (25)$$

in which Λ is a $t\times t$ positive-semidefinite diagonal matrix with diagonal elements λ_{kk} that is introduced in order to provide one with the option of weighting differently the various error vector components. It is a simple matter to show that an autoregressive coefficient vector which will render this quadratic functional a minimum must satisfy

$$X^{\dagger} Y \Lambda Y^{\dagger} X \hat{\underline{a}}^{\circ} = -X^{\dagger} Y \Lambda Y^{\dagger} \underline{x} \quad (26)$$

One then simply solves this consistent system of p linear equations in the p unknown autoregressive coefficients to obtain an estimate for the denominator of the ARMA model.

Moving-Average Coefficient Estimation

There exist several procedures for estimating the ARMA model's moving average coefficients. We shall now briefly describe two procedures which have provided satisfactory performance and in a sense complement one another.

(i) c_k Method

The procedure which has provided the best frequency resolution behavior is a direct adaption of the c_k method as described in Section II (see ref. [8]). In particular, using the set of autoregressive coefficient estimates as obtained from expression (26) and a suitable set of autocorrelation estimates $\hat{r}_x(n)$ for $n = 0,1,\ldots,\max(q,p)$, one computes the $\hat{c}_k$ coefficients using expression (13). These coefficients are then used to achieve the desired ARMA spectral estimate when incorporated into relationship (14a) and ultimately relationship (15). Although providing an excellent frequency resolution behavior, this procedure suffers the drawback of not having a guaranteed nonnegative definite spectral density estimate[3]. It is with this in mind that the following procedure was evolved.

(ii) Smoothed Periodogram Method

In the smoothed periodogram approach, one first computes the so-called "residual time-series elements according to the relationship (see ref.[9])

$$\varepsilon(n) = x(n) + \sum_{k=1}^{p} \hat{a}_k^{\circ} x(n-k) \qquad \text{for } p < n \le N \quad (27)$$

in which the $\hat{a}_k^{\circ}$ autoregressive coefficients as obtained by solving expression (26) are incorporated. From this relationship, it is apparent that the following spectral density expression holds

$$S_x(e^{j\omega}) = \frac{S_{\varepsilon}(e^{j\omega})}{|\hat{A}_p^{\circ}(e^{j\omega})|^2} \quad (28)$$

If $S_x(e^{j\omega})$ is to correspond to an ARMA spectral model of order (p,q), it is clear that a q^{th} order MA spectral estimate for the residual spectral density $S_{\varepsilon}(e^{j\omega})$ must be obtained and then substituted into relationship (28). The smoothed periodogram has been found to be a useful tool for this purpose.

In the smoothed periodogram method, one first partitions the computed residual elements (27) into

---

[1] A more generalized version of this estimation scheme can be obtained by substituting the integer k for p wherever p appears in relationship (19)-(21). For ease of presentation, k is here restricted to be p.

[2] A little thought will convince oneself that this same conclusion will be reached if both q and p are at least equal to the numerator and denominator orders, respectively, of the underlying ARMA time series.

[3] This shortcoming may be superficially avoided by taking the absolute value of the spectral estimate.

L segments each of length q+1 as specified by

$$\varepsilon_k(n) = \varepsilon(n+p+1+kd) \qquad \begin{matrix} 0\le n\le q \\ 0\le k\le L-1 \end{matrix} \tag{29}$$

where "d" is a positive integer which specifies the time shift between adjacent segments. These individual segments will overlap if $d\le q$ and will perfectly partition the residual sequence when d=q+1. In order to include only computed elements, the relevant parameters must be selected so that $q+p+1+(L-1)d\le N$. Next the periodogram for each of thee L segments is taken and these are averaged to obtain the desired q^{th} order smoothed periodogram, that is

$$\hat{S}_\varepsilon(e^{j\omega}) = \frac{1}{L}\sum_{k=0}^{L-1}\left\{\frac{1}{q+1}\left|\sum_{n=0}^{q} w(n)\varepsilon_k(n)e^{-j\omega n}\right|^2\right\} \tag{30}$$

where w(n) is a window sequence that is normally selected to be rectangular (i.e., w(n)=1 for $0\le n\le q$). The required ARMA spectral model is then obtained by substituting this approximation into relationship (28) thereby giving

$$\hat{S}_x(e^{j\omega}) = \frac{\hat{S}_\varepsilon(e^{j\omega})}{|\hat{A}_p^{\circ}(e^{j\omega})|^2} \tag{31}$$

It is readily shown that the smoothed periodogram procedure results in a desired nonnegative q^{th} order MA spectral density estimate. Unfortunately, its frequency resolution capability is generally not of the same quality as that of the c_k method.[4] On the other hand, the smoothed periodogram method provides more smoothly behaved spectral estimates which contain fewer spurious effects.

To summarize, the required ARMA spectral model is obtained by following the systematic procedure outlined in Table 2. The numerator dynamic estimation procedure to be used will of course depend on the particular characteristic being sought (e.g., frequency resolution, smoothness, etc.).

1. Specify values for the ARMA model's order parameter pair (p,q), the Yule-Walker equation parameter t, and, the weighting matrix's diagonal elements λ_{kk}.
2. Using the time series observations x(1),x(2),...,x(N), construct the matrices X, Y, and vector $\underline{x}$ according to relationships (19), (20), and (21), respectively.
3. Determine the model's autoregressive coefficients by solving relationship (26)
4. The numerator's dynamics are obtained by using either the (i) c_k method, or, (ii) the smoothed periodogram method.

Table 2. Basic steps of the standard high performance ARMA spectral estimation method: The Block Processing Mode.

The improved spectral estimation performance obtained in using this high performance method over contemporary ARMA techniques such as the Box-Jenkins method is, to a large extent, a consequence of selecting the integer t to be larger than the minimal value p. With the corresponding larger set of Yule-Walker equations that are thereby being approximated, it intuitively follows that the model's autoregressive coefficients will be less sensitive to autocorrelation estimate errors which are embodied in $Y^\dagger X$ and $Y^\dagger\underline{x}$ than would be the case if t were set to p (as in the Box-Jenkins method). This anticipated improvement in spectral estimation behavior when using the high performance method has in fact been realized on a rather large number of numerical examples [6]-[9]. As we will see in part 2 this high performance method also lends itself to a particular fast adaptive implementation mode when t=p. With the two attributes of improved spectral estimation performance and computational efficiency, this new procedure promises to be an important spectral estimation tool.

It is of interest to note that when q=0 and t=p, the high performance ARMA spectral estimation method reduces to the well known AR covariance method. Moreover, upon letting t exceed p, the resultant set of expanded AR Yule-Walker equation approximations will typically result in better spectral estimates than the standard AR covariance method. To the authors knowledge, this approach has not been used in the various AR spectral estimation procedures developed to date.

IV. ORDER SELECTION

One of the important considerations when using the high performance method is that of selecting the ARMA model order pair (p,q). This selection process can be made by utilizing properties of the ARMA autocorrelation matrix as outlined in Theorem 1. In particular, one examines the column rank behavior of the autocorrelation matrix estimate

$$\hat{R}_{tp}^{q} = Y^\dagger X \tag{32}$$

that is being used in the high performance method. Upon setting q = t = p, it follows that the p×p autocorrelation matrix estimate $\hat{R}_{pp}^{p}$ will start becoming ill-conditioned when the order parameter p exceeds the time series' inherent order value (assuming that $q\le p$). Thus, the model order determination can be achieved by investigating the conditioning of the matrix $\hat{R}_{pp}^{p}$ as a function of p. As p is increased, an appropriate choice will be a value $\hat{p}$ for which there is a precipitate decrease in matrix conditioning for $p=\hat{p}+1$. This approach, as applied to the high performance method of spectral estimation, has been used successfully by Pao and Lee [13].

There exist many matrix conditioning measures which may be used for this order determination. One of the more effective measures is the normalized determinant as specified by

$$C(A) = \det(A)\Big/\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}|a_{ij}|^2} \tag{33}$$

where det(A) designates the determinant of the p×p matrix A. It is to be noted that this normalized determinant will be zero when the rank of A is less than p.

V. THE DOWN SHIFT OPERATOR

In the analysis to follow, extensive use of the down shift operator S is made. This operator downshifts by one unit the elements of the vector upon which it operates and inserts a zero into the vacated first component position. In other words, this operation takes the form

$$S\underline{x} = [0,\ x(1),\ x(2),\ .\ .\ .,\ x(N-1)]' \tag{34a}$$

where the N×1 vector being operated upon is given by

$$\underline{x} = [x(1),\ x(2),\ldots,\ x(N)]' \tag{34b}$$

[4] A similar approach shares the same attributes as does the smoothed periodogram. [12].

The prime symbol here used denotes the operation of vector transposition. It is a simple matter to show that the downshift operator has the following N×N matrix representation

$$S = [\underline{e}_2 \vdots \underline{e}_3 \vdots \cdots \vdots \underline{e}_N \vdots \underline{\theta}] \tag{35}$$

in which $\underline{\theta}$ is the N×1 zero vector and $\underline{e}_k$ designates the k^{th} standard N×1 basis vector whose components are all zero except for its k^{th} which is one. If this downshift operator were applied sequentially m times to the vector x, it is clear that a downshift of m units results, that is

$$S^m\underline{x} = [\underbrace{0,\ 0,\ \ldots,\ 0}_{\text{m zeros}},\ x(1), x(2), \ldots, x(N-m) \tag{36}$$

VI. PREWINDOW MODIFICATION

In many spectral estimation applications, it is necessary to update the ARMA model's coefficients as new time series observations become available. If this is to be achieved in real time, however, it is generally not feasible to apply the block processing implementation of the high performance method as outlined in Table 2. In Part 2 of this paper, a computationally efficient algorithm for achieving this coefficient updating is developed. In order to facilitate this real time recursive algorithm, it is necessary to slightly modify the constituent matrices X and Y, and the vector $\underline{x}$ which characterize the high performance method. These modifications provide the required algebraic structure to render the resultant modified high performance ARMA modeling method amenable to a computationally efficient recursive solution.

Although a number of modifications are possible, we shall only treat the prewindowing method in this Section.[5] In the premodification method, the $\underline{x}$ vector is modified to

$$\underline{x} = [x(1),\ x(2),\ \ldots,\ x(N)]' \tag{37}$$

while the X matrix is modified to the N×p Toeplitz type matrix

$$X = \begin{bmatrix} 0 & 0 & 0 \\ x(1) & 0 & 0 \\ x(2) & x(1) & \cdot \\ \cdot & x(2) & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & x(1) \\ \cdot & \cdot & \vdots \\ x(N-1) & x(N-2) & x(N-p) \end{bmatrix}$$

$$= \left[\underline{Sx} \vdots S^2\underline{x} \cdots \vdots S^p\underline{x}\right] \tag{38}$$

where S is the downshift operator. Finally, the Y matrix is modified to the N×t Toeplitz type matrix

$$Y = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & & \cdot \\ 0 & 0 & & \cdot \\ x(1) & 0 & & \cdot \\ x(2) & x(1) & & 0 \\ \cdot & \cdot & & x(1) \\ \cdot & \cdot & & \vdots \\ \cdot & \cdot & & \\ x(N-q-1) & x(N-q-2) & & x(N-q-t) \end{bmatrix}$$

$$= \left[S^{q+1}\underline{x} \vdots S^{q+2}\underline{x} \vdots \cdots \vdots S^{q+t}\underline{x}\right] \tag{39}$$

Upon examination of these expressions for the modified matrices X and Y, it is seen that they possess a very simple shift type structure. It is this very structure which renders the prewindowed modification amenable to a computationally efficient adaptive solution algorithm. Furthermore, it is to be noted that lower triangular p×p and p×t matrices have been added to the top of the original X and Y matrices to form the modified X and Y matrices, respectively. These augmenting lower triangular matrices are uniquely specified so as to make the modified matrices Toeplitz in structure (i.e., the elements along any diagonal are all equal) with zeros appearing in the upper right portion of each matrix. It is this specific structure which makes an efficient recursive solution possible. This method is referred to as prewindowing since the implicit assumption that $x(n) = 0$ for $n \le 0$ is being made.

If these modifications are incorporated into expression (26), a modified set of p linear equations in the p autoregressive coefficient unknowns is obtained, that is

$$X^\dagger Y \Lambda Y^\dagger X\, \underline{\hat{a}}^\circ = -X^\dagger Y \Lambda Y^\dagger \underline{x} \tag{40}$$

This system of equations represents the least-squares solution to the following statistical approximation of the first t Yule-Walker equations

$$\underline{e} = Y^\dagger X\, \underline{a} + Y\, \underline{x} \tag{41}$$

The effectiveness of this approximation can be evaluated by taking the expected value of this relationship. When the ARMA model order parameters are such that $q \le p$, this expectation is found to give

$$E\{\underline{e}(k)\} = \begin{cases} (N-q-k)\sum_{m=0}^{q+k} a_m r_x(q+k-m) + \sum_{m=q+k+1}^{p} (N-m) a_m r_x(q+k-m), & 1 \le k < p-q \\ (N-q-k)\sum_{m=0}^{p} a_m r_x(q+k-m), & p-q \le k \le t \end{cases} \tag{42a}$$

where $a_0 = 1$. This implies that the Yule-Walker equation estimate (42) is biased in nature. As the data length N increases, however, this estimate becomes asymptotically unbiased. For the ordering case $q > p$, the expectation is found to yield

$$E\{\underline{e}(k)\} = (N-q-k)\sum_{m=0}^{p} a_m r_x(q+k-m), \quad 1 \le k \le t \tag{42b}$$

which is unbiased in nature. Thus, the set of linear equation estimates(41) generally provides a satisfactory estimate for the associated Yule-Walker equations.

In order to achieve the recursive update capability as mentioned previously, it will be necessary to "restrict" the parameter t to be p. This in turn results in $Y^\dagger X$ being a p×p matrix. When this matrix is invertible, there always exists a unique autoregressive vector which will render the error vector to be zero, that is

$$Y^\dagger X\, \underline{\hat{a}}^\circ = -Y^\dagger \underline{x} \tag{43}$$

The update algorithm to be presented in Part 2, in effect, allows us to recursively obtain the solution for the N+1 data length case from the solution to the N data length case [14]. Unfortunately, the restriction of t = p also generally results in an associated decrease in spectral estimation performance (relative to t>p). Thus, in obtaining a computationally efficient update recursive algorithm, an accompanying decrease in spectral estimation performance is the price being paid. One must therefore carefully consider the ramifications of this tradeoff in any given application. It is noteworthy, however, that this performance degradation

[5] The postwindowing, and, pre & postwindowing modification methods are described in the Appendix.

diminishes as the number of time series observations N grows.

VII. GENERALIZED LEVINSON ALGORITHM

In the high performance ARMA modeling procedures presented in Sections III and VI, the model's p autoregressive coefficients were obtained by solving a system of p linear equations. In the special case in which t = p and the p×p matrix $Y^{\dagger}X$ is nonsingular, this relevant system of equations (26) simplifies to

$$Y^{\dagger}X\,\underline{\hat{a}}^{\circ} = -Y^{\dagger}\underline{x} \tag{44}$$

where the entries of the matrices X and Y and the vector $\underline{x}$ are dependent on the particular form being used (i.e., unmodified, prewindowed, postwindowed, etc.)

If standard matrix inversion techniques such as the Cholesky decomposition method are used, on the order of p^3 multiplications and additions are required to compute the solution to relationship (44). These standard techniques are therefore said to possess a computational complexity of $O(p^3)$. For relatively large values of p, this can result in an undesirable computational burden. On the other hand, if the p×p matrix $Y^{\dagger}X$ has a near Toeplitz structure, it is possible to utilize the generalized Levinson algorithm to obtain the required solution using far fewer computations [10], [11]. Since the matrix $Y^{\dagger}X$ is being used to approximate the Toeplitz matrix R^{q}_{pp}, there is good reason to anticipate that $Y^{\dagger}X$ might possess this structural feature.

To measure the degree to which $Y^{\dagger}X$ is Toeplitz in structure, it is necessary to introduce the concept of displacement rank. The displacement rank $\alpha(A)$ of the p×p matrix A is formally given by

$$\alpha(A) = \min[\alpha_{-}(A),\ \alpha_{+}(A)] \tag{45a}$$

where

$$\alpha_{-}(A) = \text{rank}[A - SAS'] \tag{45b}$$

$$\alpha_{+}(A) = \text{rank}[A - S'AS] \tag{45c}$$

in which S is the aforementioned down shift operator (35). When the matrix A is Toeplitz, it is readily shown that its displacement rank is two (or less). Thus, a matrix whose displacement rank is near two is said to be close to Toeplitz in structure and therefore amenable to efficient inversion using the generalized Levinson algorithm.

If the displacement rank of the p×p matrix $Y^{\dagger}X$ is α, it has been shown that one can use the generalized Levinson algorithm to solve expression (44) with a corresponding computational complexity of $O(\alpha p^2)$[6]. If α is sufficiently smaller than p, a significant computational savings can be thereby realized relative to standard matrix inversion routines. Fortunately, the displacement rank of $Y^{\dagger}X$ is adequately small for the unmodified high performance ARMA modeling method and its prewindowed version (as well as the postwindowed and pre & postwindowed versions). This is a direct consequence of the fact that the columns of matrices X and Y are simply shifted versions on one another. One may readily show that the displacement rank of matrix $Y^{\dagger}X$ for each of the high performance methods is as shown in Table 3. Since these displacement ranks are so small, it is clear that the generalized Levinson algorithm may be advantageously used for solving the linear system of equations (44).

[6] As a byproduct of this solution procedure, the optimal autoregressive coefficient vectors for all ARMA models of autoregressive order k are obtained for $1 \le k \le p$.

Method	Displacement Rank $\alpha(Y^{\dagger}X)$
Standard	4
Prewindow	3
Postwindow	3
Pre & Postwindow	2

Table 3: Displacement rank of the matrix $Y^{\dagger}X$ for the various high performance ARMA methods

When the parameter t is allowed to increase beyond p so as to obtain an improved spectral estimation performance, the displacement rank of each of the methods spelled out in Table 3 increases. It is readily shown that for t>p the displacement rank increases to $\alpha^2(Y^{\dagger}X)$ in all cases. For example, the displacement rank of the t×p matrix $Y^{\dagger}X$ for the standard procedure increases to $(4)^2 = 16$ and so forth. For excessively large values of p, it would then be advantageous to use the generalized Levinson algorithm to solve relationship (44) when case t>p. The computational complexity thereby obtained would be on the order of $\alpha^2 p^2$.

VIII. NUMERICAL EXAMPLE

The unmodified ARMA modeling method of spectral estimation, as presented in Section III, has been found to possess a significantly superior performance when compared to such contemporary alternatives as the periodogram, maximum entropy, and, the Box-Jenkins methods when applied to "narrow" band time series (i.e., sums of sinusoids in white noise [6]-[9] and [13]). With this in mind, the effectiveness of both the unmodified and modified ARMA modeling procedures will now be examined for a "moderately wide band" time series. In particular, we shall treat the time series as recently considered by Bruzzone and Kaveh [15]. Specifically, their ARMA time series of order (4,4) is characterized by

$$x_k = x_k^1 + x_k^2 + 0.5\varepsilon_k \tag{46a}$$

where the individual time series x_k^1 and x_k^2 are generated according to

$$\begin{aligned} x_k^1 &= 0.4x_{k-1}^1 - 0.93x_{k-2}^1 + \varepsilon_k^1 \\ x_k^2 &= -0.5x_{k-1}^2 - 0.93x_{k-2}^2 + \varepsilon_k^2 \end{aligned} \tag{46b}$$

in which the ε_k, ε_k^1, and ε_k^2 are uncorrelated Gaussian random variables with zero mean and unit variance. It then follows that the spectral density characterizing time series (46) is given by

$$S_x(\omega) = \left|1 - 0.4e^{-j\omega} + 0.93e^{-j2\omega}\right|^{-2} + \left|1 + 0.5e^{-j\omega} + 0.93e^{-j2\omega}\right|^{-2} + 0.25 \tag{47}$$

Using the time series description (46), twenty different sampled sequences each of length 64 were generated. These twenty observation sets were then used to test various spectral estimation methods. In Figure 1, the twenty superimposed plots of the ARMA model spectral estimates of order (4,4) obtained using the first iterate of the Box-Jenkins method, and, this paper's unmodified method with $\lambda_{kk} = (0.95)^{k-1}$ and selections of t = 4, 8, and, 20 are shown. For comparison purposes, the ideal spectrum (47) is also shown. From these plots, two observations may be made: (i) the unmodified method with t = 4 yields a marginally better spectral estimate than the Box-Jenkins method, and, (ii) the unmodified spectral estimates improve

significantly as t is increased from the minimal value 4. This latter observation is most noteworthy and indicates that the incorporation of more than the minimal number of Yule-Walker equations for determining the ARMA model's autoregressive coefficients has the anticipated effect of significantly improving spectral estimation performance.

Next, the modification methods developed in Section V and the appendix were applied to these twenty different sampled sequences of length 64 to obtain ARMA model spectral estimates of order (4,4). The resultant spectra are shown in Figure 2 where it is apparent that only "a modest" degradation in spectral estimation performance has accrued due to the transient effects introduced by the modified methods. This is indeed welcomed news given the ability to implement these modified methods with exceptionally fast algorithms. It is to be noted that the "postmodified", and the "pre & postmodified" methods are identical in this example.

As a final example, twenty different sampled sequences each of length 200 were generated according to expression (46). With this longer data length, it was anticipated that an improvement in spectral estimation performance would result. A marked improvement is in fact realized as is made evident from Figure 3 where the ARMA model spectral estimates of order (4,4) are shown for the Box-Jenkins method and the unmodified method for selections of t = 4, 8, and 20.

IX. CONCLUSION

A computationally efficient closed form method of ARMA spectral estimation has been presented. It is predicated on the approximation of a set of Yule-Walker equation estimates which are generated from a given set of time series observations. The ARMA model's autoregressive coefficients are determined by solving a consistent system of linear equations. The displacement rank of the matrix corresponding to these equations is four thereby indicating that an efficient algorithmic solution procedure is possible.

The spectral estimation performance of this ARMA modeling procedure has been empirically found to exceed that of such counterparts as the maximum entropy and Box-Jenkins methods (e.g., see refs. [6]-[9] & [13]). This behavior is to a large extent, a consequence of the fact that more than the minimal number of Yule-Walker equation estimates are being approximated to obtain the resultant ARMA model parameters.

In order to achieve an improved computational efficiency, a prewindowed modification of the proposed ARMA model spectral method was next introduced. The spectral estimation performance of this prewindowed version has been found to be of high quality for moderate data lengths. As we will see in Part 2, this prewindowed method may be implemented by an adaptive update algorithm whose computational efficiency is comparable to that achieved by recently developed LMS fast algorithms.

X. REFERENCES

[1] R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communication engineering", New York: Dover, 1958.

[2] A. V. Oppenheim and R. W. Schafer, DIGITAL SIGNAL PROCESSING, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1975.

[3] G. Box and G. Jenkins, TIME SERIES ANALYSIS: FORECASTING AND CONTROL (revised edition) San Francisco: Holden-Day, 1976.

[4] P. R. Gutowski, E. A. Robinson, and S. Treitel, "Spectral estimation, fact or fiction", IEEE Transactions on Geoscience Electronics, vol. GE-16, April 1978, pp.80-84.

[5] S. A. Tretter and K. Steiglitz, "Power spectrum identification in terms of rational models", IEEE Transactions on Automatic Control, vol. AC-12, April, 1967, pp. 185-188.

[6] J. A. Cadzow, "ARMA spectral estimation: an efficient closed-form procedure", Proceedings of the RADC Spectrum Estimation Workshop, October 3-5, 1979, pp. 81-97.

[7] J. A. Cadzow, "ARMA spectral estimation: a model equation error procedure", IEEE 1980 International Conference on Acoustic, Speech and Signal Processing, April, 1980, pp. 598-602.

[8] J. A. Cadzow, "High performance spectral estimation - A new ARMA method", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.ASSP-28, No. 5, October, 1980, pp. 524-529.

[9] J. A. Cadzow, "Autoregressive moving average spectral estimation: a model equation error procedure", IEEE Transactions on Geoscience and Remote Sensing, vol.GE-19, No.1, Jan.1981, pp.24-28.

[10] T. Kailath, S. Y. Kung and M. Morf, "Displacement ranks of matrices and linear equations", Journal of Mathematical Analysis and Applications, vol. 68, No. 2, April, 1979, pp. 395-407.

[11] B. Friedlander, M. Morf, T. Kailath and L. Ljung, "New inversion formula for matrices classified in terms of their distance from Toeplitz matrices", Linear Algebra and its Applications, vol. 27, 1979, pp. 31-60.

[12] S. Kay, "A new ARMA spectral estimator", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, No. 5, Oct. 1980, pp. 585-588.

[13] Y. Pao and D. T. Lee, "Additional results on Cadzow's ARMA method for spectrum estimation", Proceedings IEEE International Conference on Acoustic, Speech, and Signal Processing, vol. 2, March 31, 1981, pp. 480-483.

[14] K. Ogino, "Computationally fast algorithms for ARMA spectral estimation", Ph.D. dissertation, Virginia Polytechnic Institute & State University, June 1981.

[15] S. Bruzzone and M. Kaveh, "On some suboptimum ARMA spectral estimators", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, Dec. 1980, pp. 753-755.

APPENDIX

I. Postwindow Modification

Following a similar procedure as employed in Section VI, the addition of an upper triangular matrix to the lower portion of the matrices specified by equations (19) and (20) yields the prewindowed matrices

$$X = \begin{bmatrix} x(p) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-p+1) \\ & \ddots & \vdots \\ \mathbf{0} & & x(N) \end{bmatrix} \qquad \text{(A1)}$$

$$Y = \begin{bmatrix} x(p-q) & \cdots & x(p-q-t+1) \\ x(p-q+1) & \cdots & x(p-q-t+2) \\ \vdots & & \\ x(N) & \cdots & x(N-t+1) \\ & \ddots & \\ \mathbf{0} & x(N) \cdots & x(N+p-q-t) \end{bmatrix} \qquad \text{(A2)}$$

where X and Y are recognized as being $(N \times p)$ and $(N \times t)$ Toeplitz type matrices, respectively. In a similar manner, the column vector $\underline{x}$ is modified to

$$\underline{x} = [x(p+1), \ldots\ldots, x(N), \underbrace{0, \ldots, 0}_{p \text{ zeros}}]' \qquad \text{(A3)}$$

The displacement rank of the matrix $Y^\dagger X$ is readily found to be 3.[7] A generalized Levinson procedure requiring a computational complexity of $O(3p^2)$ can then be applied for solving the system of equations

$$Y^\dagger X \, \underline{\hat{a}} = -Y^\dagger \underline{x} \tag{A4}$$

A more computationally efficient algorithm associated with the postwindow modification has been developed [14]. It is shown that the number of computations is reduced to (p log p) if p = q where p and q are the denominator and numerator orders of the ARMA model, respectively.

II. Pre & Postwindow Modification Method

The combination of the previously discussed prewindowed and postwindowed modification methods yields the pre & postwindow modification method. The matrices and vectors are modified in the following manner.

$$X = \begin{bmatrix} 0 & \cdots & 0 \\ x(1) & \ddots & \vdots \\ \vdots & \ddots & 0 \\ x(p) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-p) \\ & \ddots & \vdots \\ \bigcirc & & x(N) \end{bmatrix} \tag{A5}$$

$$Y = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ x(1) & \ddots & 0 \\ & \ddots & \\ x(t) & \cdots & x(1) \\ \vdots & & \vdots \\ x(N) & \cdots & x(N-t+1) \\ \vdots & & \vdots \\ \bigcirc & x(N) \cdots & x(N+p-q-t) \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} q \text{ rows} \tag{A6}$$

$$\underline{x} = [x(1), \ldots x(N), \underbrace{0, \ldots, 0}_{p \text{ zeros}}] \tag{A7}$$

where X and Y denote (N+p)×p and (N+p)×t Toeplitz type matrices, respectively, and $\underline{x}$ denotes a (N+p)×1 column vector.

It can be shown that $Y^\dagger X$ is a Toeplitz matrix. The conventional Levinson algorithm may therefore be used for solving the Toeplitz system of equations

$$Y^\dagger X \underline{a} = -Y^\dagger x \tag{A8}$$

in which the inherent computational complexity is $O(2p^2)$.[7] More recently, a fast algorithmic solution has been developed which significantly reduces this computational complexity.

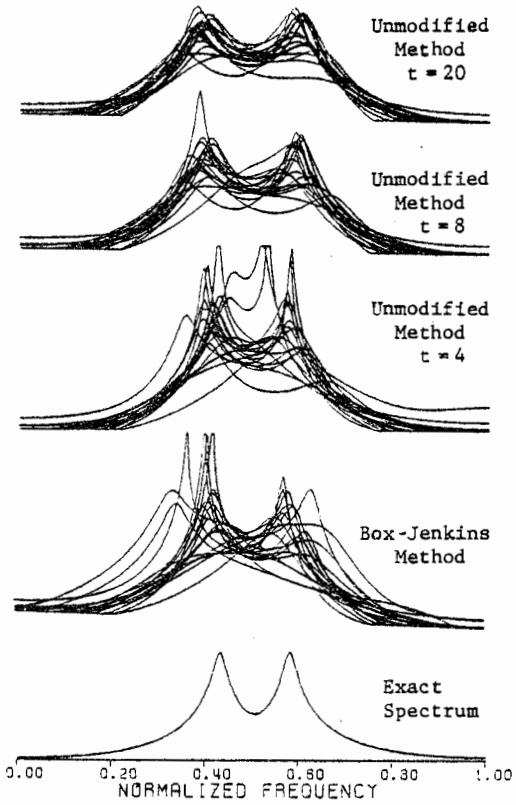

Figure 1: Spectral estimates of order (4,4) by the Box-Jenkins method and by the High Performance method using various values for t. N = 64 data points for each estimate.

[7] The parameter t is here taken to equal p.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 9	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): An Adaptive ARMA Spectral Estimator: Part 1		5. TYPE OF REPORT & PERIOD COVERED: Conference Paper
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): J. A. Cadzow & R. L. Moses		8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0303
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Virginia Tech, Department of Electrical Engineering, Blacksburg, VA 24061		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research (Code 436), Statistics and Probability Program, Arlington, VA 22217		12. REPORT DATE: June 1981 13. NUMBER OF PAGES: 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ARMA Spectral Estimate, Block Processing Algorithm, Yule-Walker Equations, Rational Time Series Modeling, Prewindowed Modification.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A novel method for generating a rational ARMA model of a time series is developed. It is predicated on approximating the underlying Yule-Walker equations and provides an excellent spectral estimation performance. Moreover, its spectral estimation performance typically exceeds that of contemporary methods.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AN ADAPTIVE ARMA SPECTRAL ESTIMATOR: PART 2*

James A. Cadzow and Randolph L. Moses
Department of Electrical and Computer Engineering
Arizona State University
Tempe, Arizona 85281

Abstract - This paper presents a recursive algorithmic implementation of the prewindowed high performance method of ARMA spectral modeling as described in Part 1. This algorithm provides updates of the ARMA models optimal autoregressive coefficients (in actuality prediction errors) as each new data point becomes available. The algorithm is computationally fast in the sense that it requires O(p) multiplications and additions for each update. It is shown that this fast recursive algorithm may be implemented using a lattice filter arrangement, and it therefore exhibits several of the "nice" properties associated with lattice type algorithms such as numerical robustness and good convergence properties.

I. INTRODUCTION

In Part 1 of this paper we described an algorithm for obtaining an estimate of the power spectral density associated with a given time series $\{x(n)\}$. In particular, the following ARMA (p,q) spectral density model was hypothesized

$$S_x(\omega) = \left| \frac{b_0 + b_1 e^{-j\omega} + \ldots + b_q e^{-jq\omega}}{1 + a_1 e^{-j\omega} + \ldots + a_p e^{-jp\omega}} \right|^2 \qquad (1)$$

A closed form algorithm for estimating this model's a_i and b_j parameters was then developed in which the finite set of time series observations

$$x(1), x(2), \ldots, x(n) \qquad (2)$$

were used in the parameter selection process. This so called "high performance" algorithm operates on a data block of length n to obtain the model's coefficients in a single computational effort. It is therefore called a block processing algorithm.

There are many situations, however, in which a block processing algorithm for spectral estimation is not an appropriate tool. In a variety of applications, the data measurement is an ongoing process and it is therefore desirable to recursively update the autoregressive and moving average parameters as each new data point becomes available. This capability is of particular importance in those cases where one wishes to adaptively model the spectrum of a long, ongoing time series. Algorithms with this recursive updating capability are called "recursive algorithms".

Recently, several fast recursive spectral estimation algorithms have been developed [1]-[7]. Most of these algorithms are based on so called least-squares AR spectral estimation methods [8], although a few ARMA algorithms have also been developed [4] & [6]. These algorithms seek to minimize a prediction error vector in order to obtain the desired spectral estimates.

In this paper we develop a more effective ARMA algorithm that efficiently computes the optimal autoregressive coefficients by recursively updating a set of prediction error elements as each new data point is observed. This recursive algorithm is based on the prewindowed high performance method as described in Part 1, and, therefore is predicated on the approximation of the ARMA model's underlying Yule-Walker equations. Although similar algorithms may be derived for the unmodified and for the two other modified versions of the high performance method, the recursive algorithm based on the prewindow version is a bit easier to derive and is characterized by a "fast start-up" capability in that spectral estimates are possible with as few as two data points. The recursive algorithm herein presented is computationally efficient in the sense that O(p) multiplications and additions are required to update the "necessary" parameters as each new data point is observed. This paper's recursive algorithm was originally developed in [9]. A more straightforward derivation is herein presented which provides a greater degree of insight. In addition, a lattice filter implementation of this algorithm is developed. Moreover, because of this ladder-type implementation, this algorithm is characterized by several other nice properties associated with ladder algorithms such as numerical robustness and good convergence properties [10], [11].

II. THE PREDICTION ERROR VECTORS

The recursive update equations herein presented do not explicitly update the ARMA model's autoregressive coefficients in obtaining optimal updated spectral estimates. Instead, a set of "equivalent" parameters known as prediction errors are updated. In this section, we discuss the relationship between the prediction errors and autoregressive coefficients.

As outlined in Part 1 of this paper, the optimal p^{th} order set of autoregressive coefficients for the prewindowed version of the high performance method are obtained by solving the following system of p linear equations in p unknowns

$$[Y_{n,p}^{\dagger} \; X_{n,p}]\underline{a}_p + Y_{n,p}^{\dagger}\,\underline{x}_n = \underline{\theta} \qquad (3)$$

where

$$\underline{x}_n = [x(1), x(2), \ldots, x(n)]' \qquad (4a)$$

$$\underline{y}_n = S^q \underline{x}_n \qquad (4b)$$

$$X_{n,p} = [S\underline{x}_n : S^2\underline{x}_n : \ldots : S^p\underline{x}_n] \qquad (4c)$$

$$Y_{n,p} = [S\underline{y}_n : S^2\underline{y}_n : \ldots : S^p\underline{y}_n] \qquad (4d)$$

in which $\underline{\theta}$ is the zero vector and S is the down shift operator.[1] Here the dagger symbol (†) denotes complex conjugate transposition and the prime symbol (') denotes transposition. The subscripts p and n explicitly indicate that the denominator order of the spectral model of equation (1) is p and that n data points are available. Whenever this explicit information is not needed, we will use $\underline{x}$, $\underline{y}$, X, and Y in place of $\underline{x}_n$, $\underline{y}_n$, $X_{n,p}$, and $Y_{n,p}$, respectively.

It is recalled from Part 1, that the high performance ARMA modeling approach is predicated on approximating t

*This work was supported in part by the Office of Naval Research, Statistics and Probability Program under Contract N00014-80-C-0303.

[1] In this paper, matrices are denoted by capital letters (e.g. X), vectors are denoted by underlined lower case English letters (e.g. $\underline{x}$) and scalars are denoted by lower case Greek letters (e.g. α). Moreover, the down shift operator S is defined by

$$S\underline{x}_n = [0, x(1), x(2), \ldots, x(n-1)]'$$

Yule-Walker equations where $t \geq p$. Upon examination of expression (3), it is apparent that we have here restricted $t = p$. This restriction is required in order to facilitate the development of the fast recursive algorithm. Unfortunately, by requiring $t = p$, the spectral estimation performance suffers in comparison to that achieved with larger values of t. As the data length n increases, however, this performance degradation diminishes and typically is of an insignificant nature. This is indeed fortunate since it is precisely for long data length cases that the recursive algorithm would most likely be utilized.

Under the assumption that $Y^\dagger X$ is nonsingular, the optimal autoregressive coefficient vector which satisfies expression (3) is given by

$$\underline{a}_p^\circ = -[Y_{n,p}^\dagger X_{n,p}]^{-1} Y_{n,p}^\dagger \underline{x}_n \tag{5}$$

In what is to follow, it is beneficial to interpret this autocorrelation coefficient selection procedure from a prediction error viewpoint. Namely, we may reformulate expression (3) as

$$Y_{n,p}^\dagger \underline{f}_{p,n}^x = \underline{0} \tag{6}$$

in which $\underline{f}_{p,n}^x$ is the so-called forward prediction error as specified by

$$\underline{f}_{p,n}^x = \underline{x}_n + X_{n,p}\, \underline{a}_p \tag{7}$$

It is referred to as the forward prediction error since its k^{th} component can be interpreted as being the error resulting from a prediction of the element x(k) by a linear combination of the p most recent time series elements x(k-1), x(k-2),. . ., x(k-p).

The optimal autoregressive coefficient vector (5) can be then associated with an auxiliary minimization problem involving the prediction error vector. Namely, it is readily shown that this optimal vector minimizes the following quadratic functional

$$g(\underline{a}_p) = \underline{f}_{p,n}^{x\dagger} W \underline{f}_{p,n}^x \tag{8}$$

where W is the $n \times n$ positive semidefinite matrix specified by

$$W = Y_{n,p}\, Y_{n,p}^\dagger \tag{9}$$

To reinforce this prediction error interpretation, let us define the following estimate of vector $\underline{x}_n$

$$\hat{\underline{x}}_n = -X_{n,p}\, \underline{a}_p \tag{10}$$

which in turn generates the forward predicting error

$$\underline{f}_{p,n}^x = \underline{x}_n - \hat{\underline{x}}_n \tag{11}$$

Upon substitution of expression (5) into (10) the optimum forward prediction error vector is given by

$$\hat{\underline{x}}_n^\circ = X_{n,p}[Y_{n,p}^\dagger X_{n,p}]^{-1} Y_{n,p}^\dagger \underline{x}_n$$

$$= P_{XY}\, \underline{x}_n \tag{12a}$$

while the minimizing forward prediction error for this selection becomes

$$\underline{\hat{f}}_{p,n}^{x\,\circ} = \left(I - X_{n,p}[Y_{n,p}^\dagger X_{n,p}]^{-1} Y_{n,p}^\dagger\right) \underline{x}_n \tag{12b}$$

$$= P_{XY}^c\, \underline{x}_n$$

We have here used the compact matrix product representations

$$P_{XY} = X_{n,p}[Y_{n,p}^T X_{n,p}]^{-1} Y_{n,p}^\dagger \tag{13a}$$

$$P_{XY}^c = I - P_{XY} \tag{13b}$$

Since we are only interested in the optimal $\hat{\underline{x}}_n$ and $\underline{f}_{p,n}^x$, we will drop the "°" symbol and assume that $\hat{\underline{x}}_n$ and $\underline{f}_{p,n}^x$ are the optimal ones as given by equation (12)

We may also define the delayed backward prediction error vector for $\underline{x}_n$ by [2]

$$\underline{d}_{p,n}^x = S^{p+1}\underline{x}_n + X_{n,p}\, \underline{\tilde{a}}_p \tag{14}$$

It can be seen that the k^{th} row of equation (14) represents a prediction of x(k-p-1) by a linear combination of the p most immediate future values x(k-p), x(k-p+1),. . .,x(k-1). The resulting error in this backward prediction is $d_{p,n}^x(k)$. In this case the optimum $\underline{\tilde{a}}_p$ vector is the one that minimizes the quadratic function

$$\tilde{g}(\underline{\tilde{a}}_p) = [\underline{d}_{p,n}^x]^\dagger W \underline{d}_{p,n}^x \tag{15}$$

where W is defined in equation (9)

In this case the optimal $\tilde{a}$ is given by

$$\underline{\tilde{a}}_p^\circ = -[Y_{n,p}^\dagger X_{n,p}]^{-1} Y_{n,p}^\dagger (S^{p+1}\underline{x}_n) \tag{16}$$

In a similar manner to the forward prediction error case the optimal estimate of $S^{p+1}\underline{x}_n$ is specified by

$$\widehat{S^{p+1}\underline{x}_n}^\circ = -X_{n,p}\, \underline{\tilde{a}}_p^\circ = P_{XY}(S^{p+1}\underline{x}_n) \tag{17a}$$

and the optimal delayed backward prediction error vector is

$$\underline{d}_{p,n}^{x\,\circ} = S^{p+1}\underline{x}_n - \widehat{S^{p+1}\underline{x}_n}^\circ = P_{XY}^c(S^{p+1}\underline{x}_n) \tag{17b}$$

where P_{XY} and P_{XY}^c are given by equation (13). Furthermore, it can be shown that the optimal $\underline{a}_p$ as given by equation (16) also arises by approximating p Yule-Walker equations in a manner similar to the approximation given by equation (3) for $\underline{a}_p$.

It is clear that the forward prediction error vector $\underline{f}_{p,n}^x$ and the autoregressive coefficient vector $\underline{a}_p$ are interchangeable in the sense that one can always be found from the other using equation (7). Similarly $\underline{d}_{p,n}^x$ and $\underline{\tilde{a}}_p$ are interchangeable since one can always be found from the other using equation (14). It is also true that the 2p elements of $\underline{a}_p$ and $\underline{\tilde{a}}_p$ are interchangeable with the 2p elements $f_{1,n}^x(n)$, $f_{2,n}^x(n)$,..., $f_{p,n}^x(n)$ and $d_{1,n}^x(n)$, $d_{2,n}^x(n)$, ..., $d_{p,n}^x(n)$ (i.e., the n^{th} elements of the 2p prediction error vectors $\underline{f}_{1,n}^x$, $\underline{f}_{2,n}^x$,..., $\underline{f}_{p,n}^x$ and $\underline{d}_{1,n}^x$, $\underline{d}_{2,n}^x$,...,$\underline{d}_{p,n}^x$). We will show this last fact in Section VIII, where we will also see that the prediction errors lead to a lattice filter structure which is related to the autoregressive coefficient vectors.

In the fast recursive algorithm, the autoregressive coefficient vectors $\underline{a}_p$ and $\underline{\tilde{a}}_p$ are not directly updated. Instead, the prediction error elements $f_{1,n}^x(n)$,..., $f_{p,n}^x(n)$, and $d_{1,n}^x(n)$,..., $d_{p,n}^x(n)$ are updated. Since these elements are interchangeable with the autoregressive coefficients, there is no information lost in updating only the prediction error elements. However, the prediction error elements may be updated in a computationally efficient manner, requiring O(p) multiplications and additions for the update. Moreover, as we shall see later, the 2p prediction error elements enable us to find all of the $\underline{a}_m$ and $\underline{\tilde{a}}_m$ vectors for ARMA denominator orders from 1 to p. It is for these reasons that we choose to update the

[2] We use the term "delayed" because although the subscript n appears, x(n) is never used in (14). The undelayed backward prediction error vector will be discussed in a later section.

prediction error elements.

We may also obtain additional autoregressive coefficient estimates similar to $\underline{a}$ and $\tilde{\underline{a}}$ by considering the prediction error vectors associated with the vector $\underline{y}_n$ given in equation (4b). Specifically, the optimum forward prediction error corresponding to $\underline{y}_n$ is defined as

$$\underline{f}^{y}_{p,n}{}^{\circ} = \underline{y}_n + Y_{n,p}\,\underline{c}_p^{\circ} \tag{18a}$$

where

$$\underline{c}_p^{\circ} = -[X_{n,p}^{\dagger} Y_{n,p}]^{-1} X_{n,p}^{\dagger}\,\underline{y}_n \tag{18b}$$

and $X_{n,p}$ $Y_{n,p}$ and $\underline{y}_n$ are defined in equation (4). As in earlier cases, $\underline{c}_p^{\circ}$ can be found by approximating p Yule-Walker equations or, equivalently, by minimizing the quadratic functional

$$h(\underline{c}_p) = [\underline{f}^{y}_{p,n}]^{\dagger}[X_{n,p}\, X_{n,p}^{\dagger}][\underline{f}^{y}_{p,n}] \tag{19}$$

Corresponding to this optimal autoregressive coefficient vector $\underline{c}_p^{\circ}$ we may also define the estimate

$$\hat{\underline{y}}_n^{\circ} = -Y_{n,p}\,\underline{c}_p^{\circ} = Y_{n,p}[X_{n,p}^{\dagger} Y_{n,p}]^{-1} X_{n,p}^{\dagger}\,\underline{y}_n$$
$$= P_{YX}\,\underline{y}_n \tag{20}$$

which, in turn, gives rise to the optimum error vector

$$\underline{f}^{y}_{p,n}{}^{\circ} = P^{c}_{YX}\,\underline{y}_n$$

where P_{YX} and P^{c}_{YX} are defined as in equation (13)

Finally, the optimal delayed backward prediction error vector for $\underline{y}_n$ is defined by

$$\underline{d}^{y}_{p,n}{}^{\circ} = S^{p+1}\underline{y}_n + Y_{n,p}\,\tilde{\underline{c}}_p^{\circ} \tag{21a}$$

where

$$\tilde{\underline{c}}_p^{\circ} = -[X_{n,p}^{\dagger} Y_{n,p}]^{-1} X_{n,p}^{\dagger}(S^{p+1}\underline{y}_n) \tag{21b}$$

Again, we can define the predicted value of $S^{p+1}\underline{y}$ by

$$\widehat{S^{p+1}\underline{y}}_n^{\circ} = Y_{n,p}[X_{n,p}^{\dagger} Y_{n,p}]^{-1} X_{n,p}^{\dagger}(S^{p+1}\underline{y}) = P_{YX}(S^{p+1}\underline{y}) \tag{22}$$

and it follows that

$$\underline{d}^{y}_{p,n}{}^{\circ} = P^{c}_{YX}\,(S^{p+1}\underline{y}_n) \tag{23}$$

Just as for the $\underline{x}_n$ vector, the 2_p entities $\underline{f}^{y}_{1,n}(n),\ldots,\underline{f}^{y}_{p,n}(n)$, and $\underline{d}^{y}_{1,n}(n)\ldots,\underline{d}^{y}_{p,n}(n)$ can be efficiently updated and enable us to determine the optimal $\underline{c}_m$ and $\tilde{\underline{c}}_m$ coefficient vectors for all ARMA model denominator orders from 1 to p.[3]

III. THE HILBERT SPACE SETTING

The problem of recursively updating the prediction error vectors in the fast algorithm can be more easily understood by casting the problem in a Hilbert space setting. Consider the n dimensional complex Euclidian space

$$H = C^{n} = C \times C \times \ldots \times C \tag{24}$$

with the standard vector inner product defined by

$$\langle \underline{x},\underline{y} \rangle = \underline{x}^{\dagger}\underline{y} = \sum_{i=1}^{n} x^{*}(i)\, y(i) \tag{25}$$

[3] Throughout the remainder of the paper, the "°" symbol will be dropped and the prediction error vectors $\underline{f}^{x}$, $\underline{d}^{x}$, $\underline{f}^{y}$, and $\underline{d}^{y}$ are assumed to be the optimal ones.

We note that the n×1 vectors $\underline{x}_n$, $S^{m}\underline{x}_n$, $\underline{y}_n$, and $S^{m}\underline{y}_n$ are all elements of H. Moreover, the p columns of matrix $X_{n,p}$ are also elements of H. The set of all linear combinations of these p elements is a subspace of H. which we denote by M_X. Similarly, M_Y is the subspace spanned by the p columns of $Y_{n,p}$.

Let us now consider the forward prediction of $\underline{x}_n$. From equation (12a) we see that $\underline{x}_n$ is formed by a matrix multiplication involving $\underline{x}_n$. The matrix P_{XY} is seen to be a linear operator on the Hilbert space H. It is apparent that P_{XY} maps elements of H into elements in the subspace M_X, that is

$$P_{XY}\colon H \rightarrow M_X \tag{26}$$

Also, it is evident from equation (13) that $P_{XY}{}^{2} = P_{XY}$ so that the operator P_{XY} is a projection operator onto the subspace M_X. In general, P_{XY} is not the orthogonal projection operator onto subspace M_X. Instead, the associated direction of projection is determined by the matrix $Y_{n,p}$. It can be seen from equation (6) that the direction of projection of P_{XY} is orthogonal to M_Y. Thus, P_{XY} is the projection operator onto the subspace M_X along $M_Y{}^{\perp}$ (the orthogonal complement of M_Y).

With these thoughts in mind, we can provide a simple geometric interpretation to the four error vectors described in the last section. In particular, the geometric relationship between $\underline{x}_n$, $\hat{\underline{x}}_n$, and $\underline{f}^{x}_{p,n}$ is depicted in Figure 1.

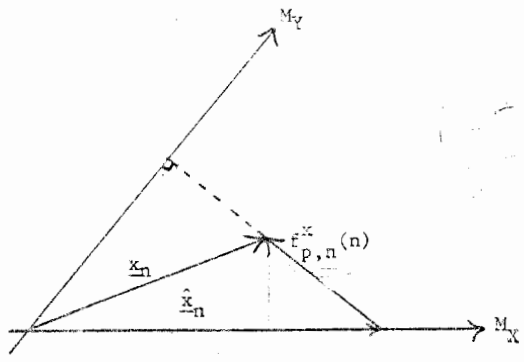

Figure 1: Geometric Relationship Between $\underline{x}_n$, $\hat{\underline{x}}_n$, and the optimal prediction error is $\underline{f}^{x}_{p,n}$

The vector $\hat{\underline{x}}_n$ is seen to be that projection of $\underline{x}_n$ onto M_X that is orthogonal to M_Y. We note from Figure 1 that

$$\underline{f}^{x}_{p,n} \perp M_Y \tag{27a}$$

or, equivalently, that

$$\langle \underline{f}^{x}_{p,n},\, S^{m}\underline{y}_n \rangle = 0, \quad m = 1,2,\ldots,p \tag{27b}$$

The geometric relationships for $\underline{d}^{x}_{p,n}$, $\underline{f}^{y}_{p,n}$, and $\underline{d}^{y}_{p,n}$ are similar to Figure 1.

Since P_{XY} and P_{YX} are projection operators, so are their complements P^{c}_{XY} and P^{c}_{YX}. It follows that

$$(P^{c}_{XY})^{2} = P^{c}_{XY} \tag{28a}$$

$$(P^{c}_{YX})^{2} = P^{c}_{YX} \tag{28b}$$

We can also see from equation (13) that P_{XY} and P_{YX} are strongly related, namely

$$P_{XY} = [P_{YX}]^{\dagger} \tag{29a}$$

and $$P^{c}_{XY} = [P^{c}_{YX}]^{\dagger} \tag{29b}$$

In addition to the four prediction error vectors, there are four inner products that are useful in deriving the fast algorithm. These complex-valued scalars are defined as:

$$\sigma_{p,n} \triangleq [\underline{f}^{x}_{p,n}]^{\dagger}[\underline{d}^{y}_{p,n}] = [\underline{x}_n]^{\dagger} P^{c}_{XY}[S^{p+1}\underline{y}_n] \tag{30}$$

$$\tau_{p,n} \triangleq [\underline{d}^x_{p,n}]^\dagger [\underline{f}^y_{p,n}] = [S^{p+1}\underline{x}_n]^\dagger P^c_{XY}[\underline{y}_n] \quad (31)$$

$$\mu_{p,n} \triangleq [\underline{f}^x_{p,n}]^\dagger [\underline{f}^y_{p,n}] = [\underline{x}_n]^\dagger P^c_{XY}[\underline{y}_n] \quad (32)$$

$$\nu_{p,n} \triangleq [\underline{d}^x_{p,n}]^\dagger [\underline{d}^y_{p,n}] = [S^{p+1}\underline{x}_n]^\dagger P^c_{XY}[S^{p+1}\underline{y}_n] \quad (33)$$

IV. THE PROJECTION OPERATOR THEOREM

From the results of the last section it is apparent that the various prediction error vectors and scalars are all described by the operators P_{XY} and P_{YX}. As a new data point x(n+1) becomes available, we desire to update the prediction errors vectors and scalars in a computationally efficient manner. Because the operators P_{XY} and P_{YX} are used repeatedly and their structures change as new data points become available, we prefer to update them and then obtain updated error vectors by applying these updated projection operators. Recursive update equations for the projection operators P_{XY} and P_{YX}, are readily obtained by appealing to the following theorem.

Theorem 1. (Projection Operator Theorem) Let A and B be n×m matrices. Furthermore, consider the augmented matrices $\bar{A} = [A : \underline{a}]$ and $\bar{B} = [B : \underline{b}]$ in which $\underline{a}$ and $\underline{b}$ are n×1 vectors. If $[A^\dagger B]^{-1}$ and $[\bar{A}^\dagger \bar{B}]^{-1}$ exist, then the associated projection operator corresponding to the augmented matrices is given by

$$P^c_{\bar{A}\bar{B}} = P^c_{AB} - P^c_{AB}\,\underline{a}[\underline{b}^\dagger P^c_{AB}\,\underline{a}]^{-1}\underline{b}^\dagger P^c_{AB} \quad (34)$$

where $$P_{AB} = A[B^\dagger A]^{-1}B^\dagger \quad (35a)$$

$$P^c_{AB} = I - P_{AB} \quad (35b)$$

The theorem may be straightforwardly proven by writing $[\bar{A}^\dagger \bar{B}]$ in terms of Schur complements and performing some matrix algebra. Alternatively, the theorem may be proven using Hilbert space concepts.

Using this projection operator theorem we may now obtain all the necessary equations for the fast recursive algorithm. Two types of recursive equations are of interest. First, equations are needed that provide the m+1st order prediction error vectors in terms of the m^{th} order errors. These are called order update equations. Second, equations are needed that enable us to update the prediction errors as a new data point becomes available. These equations are referred to as the time update recursions.

These two sets of equations are derived below.

V. ORDER UPDATE RECURSIONS

In this section the order update equations for $\underline{f}^x$, $\underline{f}^y$, $\underline{d}^x$, $\underline{d}^y$, μ and ν are derived by making use of the projection operator theorem.

Consider first the error vector $\underline{f}^x_{m+1,n}$ associated with the optimum m+1st order autoregressive coefficients. Here, m can take on any value in the range 0, 1, ..., p − 1, where p is the desired autoregressive coefficient order. From equations (4) and (12b), we see that

$$\underline{f}^x_{m+1,n} = P^c_{\bar{X}\bar{Y}}\,\underline{x}_n \quad (36a)$$

where $$\bar{X} = X_{n,m+1} = [X_{n,m} : S^{m+1}\underline{x}_n] \quad (36b)$$

$$\bar{Y} = Y_{n,m+1} = [Y_{n,m} : S^{m+1}\underline{y}_n] \quad (36c)$$

Applying the projection operator theorem to (36a) with A = X, B = Y, $\underline{a} = S^{m+1}\underline{x}$, and $\underline{b} = S^{m+1}\underline{y}$, we have

$$\underline{f}^x_{m+1,n} = P^c_{XY}\,\underline{x} - P^c_{XY}(S^{m+1}\underline{x})[(S^{m+1}\underline{y})^\dagger P^c_{XY}(S^{m+1}\underline{x})]^{-1}(S^{m+1}\underline{y})^\dagger P^c_{XY}\,\underline{x}$$

$$\underline{f}^x_{m+1,n} = \underline{f}^x_{m,n} - \frac{\sigma_{m,n}}{\nu^*_{m,n}}\,\underline{d}^x_{m,n} \quad (37)$$

As mentioned earlier, to implement the recursive algorithm we only need the error element at time n, that is $f^x_{m+1,n}(n)$. From equation (37) we see that

$$f^x_{m+1,n}(n) = f^x_{m,n}(n) - \frac{\sigma^*_{m,n}}{\nu^*_{m,n}}\,d^x_{m,n}(n) \quad (38)$$

Equation (38) is the desired order update equation for f^x.

In a similar manner, the order update equation for f^y is found to be

$$\underline{f}^y_{m+1,n} = P^c_{\bar{Y}\bar{X}}\,\underline{y}_n \quad (39)$$

where $\bar{X}$ and $\bar{Y}$ are defined in equation (36). Applying the projection operator theorem yields

$$\underline{f}^y_{m+1,n} = \underline{f}^y_{m,n} - \frac{\tau_{m,n}}{\nu_{m,n}}\,\underline{d}^y_{m,n} \quad (40)$$

The n^{th} component of equation (40) is the desired order update equation for the forward y prediction error, that is

$$f^y_{m+1,n}(n) = f^y_{m,n}(n) - \frac{\tau_{m,n}}{\nu_{m,n}}\,d^y_{m,n}(n) \quad (41)$$

The order update equations for the delayed backward prediction error vectors may be similarly derived. These equations are, however, not as useful as the combined order and time update equations. The combined update equations give $\underline{d}^x_{m+1,n+1}$ and $\underline{d}^y_{m+1,n+1}$ in terms of $\underline{d}^x_{m,n}$ and $\underline{d}^y_{m,n}$. In deriving the combined order and time update equation for the delayed backward $\underline{x}_n$ estimate, we first note that from equation (17b)

$$\underline{d}^x_{m+1,n+1} = P^c_{\bar{X}\bar{Y}}(S^{m+1}\underline{x}_n) \quad (42a)$$

where $$\bar{X} = X_{n+1,m+1} = \begin{bmatrix} 0' & : & 0 \\ \cdots & : & \cdots \\ X_{n,m} & : & \underline{x}_n \end{bmatrix} \quad (42b)$$

$$\bar{Y} = Y_{n+1,m+1} = \begin{bmatrix} 0' & : & 0 \\ \cdots & : & \cdots \\ Y_{n,m} & : & \underline{y}_n \end{bmatrix} \quad (42c)$$

We now apply the projection operator theorem to equation (42a) with $A = \bar{X}$ and $B = \bar{Y}$ in equation (34). After some simple algebraic manipulation, we get

$$\underline{d}^x_{m+1,n+1} = \begin{bmatrix} 0 \\ \underline{d}^x_{m,n} \end{bmatrix} - \frac{\tau^*_{m,n}}{\mu^*_{m,n}} \begin{bmatrix} 0 \\ \underline{f}^x_{m,n} \end{bmatrix} \quad (43)$$

The n+1st component of equation (43) yields the desired update equation

$$d^x_{m+1,n+1}(n+1) = d^x_{m,n}(n) - \frac{\tau^*_{m,n}}{\mu^*_{m,n}}\,f^x_{m,n}(n) \quad (44)$$

The delayed backward time and order update equation for $\underline{y}_n$ is derived in a similar manner. The details are omitted, but it is readily shown that

$$d^y_{m+1,n+1}(n+1) = d^y_{m,n}(n) - \frac{\sigma_{m,n}}{\mu_{m,n}}\,f^y_{m,n}(n) \quad (45)$$

Finally, the order update equations for the scalars $\mu_{m,n}$ and $\nu_{m,n}$ are derived. From equation (32)

$$\mu_{m+1,n} = \underline{y}_n^\dagger P^c_{\bar{X}\bar{Y}}\,\underline{x}_n \quad (46a)$$

where $$\bar{X} = X_{n,m+1} = [X_{n,m} : (S^{m+1}\underline{x}_n)] \quad (46b)$$

$$\bar{Y} = Y_{n,m+1} = [Y_{n,m} : (S^{m+1}\underline{y}_n)] \quad (46c)$$

Applying the projection operator theorem gives the desired result

$$\mu_{m+1,n} = \underline{y}^{+}P^{c}_{XY}\,\underline{x}$$

$$-\underline{y}^{+}P^{c}_{XY}(S^{m+1}\underline{x})[(S^{m+1}\underline{y})^{+}P^{c}_{\overline{XY}}(S^{m+1}\underline{x})]^{-1}(S^{m+1}\underline{y})^{+}P^{c}_{XY}\,\underline{x}$$

$$\mu_{m+1,n} = \mu_{m,n} - \frac{\sigma_{m,n}\tau_{m,n}}{\nu^{*}_{m,n}} \tag{47}$$

Also, from equation (33)

$$\nu_{m+1,n+1} = (S^{m+1}\underline{y})^{+}P^{c}_{\overline{XY}}(S^{m+1}\underline{x}_{n}) \tag{48}$$

where $\bar{X}$ and $\bar{Y}$ are defined in equation (46). Applying the projection operator theorem yields the combined order and time update equation for ν.

$$\nu_{m+1,n+1} = \nu_{m,n} - \frac{\sigma_{m,n}\tau_{m,n}}{\mu^{*}_{m,n}} \tag{49}$$

VI. TIME UPDATE EQUATIONS

The remaining recursive equations update the forward and delayed backward errors as a new data point is obtained. For this reason these equations are called time update equations.

When a new data point becomes available, the effect on the prediction error vectors is to append a row to the bottom of their defining matrix equation [see, for example, equation (7)]. Appending a row to the matrices $X_{n,m}$ and $Y_{n,m}$ does not seem to fit in the framework of the projection operator theorem, in which columns are appended to $X_{n,m}$ and $Y_{n,m}$. It turns out, however, that we can accomplish the task of annihilating a row in the error vector matrix equation by appending to $X_{n,m}$ and $Y_{n,m}$ the $n^{\underline{th}}$ basis vector defined by

$$\underline{e}_{n} = [0 \;\; 0 \;\ldots\; 0 \;\; 1]^{t} \tag{50}$$

To see how this works, let us consider as an example the forward prediction error vector for $\underline{x}_n$. If we append the n×1 vector $\underline{e}_n$ to $X_{n,m}$ in equation (7) we have

$$\begin{bmatrix} \tilde{f}^{x}_{m,n}(1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \tilde{f}^{x}_{m,n}(n) \end{bmatrix} = \begin{bmatrix} x(1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x(n) \end{bmatrix} + \left[\begin{array}{ccc:c} 0 & \cdots & 0 & 0 \\ x(1) & & \vdots & 0 \\ \cdot & & 0 & \vdots \\ \cdot & & x(1) & \vdots \\ \cdot & & \cdot & \vdots \\ \cdot & & \cdot & 0 \\ x(n-1) & & x(n-m) & 1 \end{array}\right] \begin{bmatrix} a_m(1) \\ a_m(2) \\ \vdots \\ a_m(m) \\ \xi \end{bmatrix} \tag{51}$$

where $\tilde{f}^{x}$ is used instead of $\underline{f}^{x}$ to indicate the presence of the $\underline{e}_n$ vector. The optimal $\tilde{f}^{x}_{m,n}$ vector is given by

$$\underline{f}^{x}_{m,n} = P^{c}_{\overline{XY}}\,\underline{x}_{n} \tag{52a}$$

where

$$\bar{X} = [X_{n,m} \,\vdots\, \underline{e}_{n}] \tag{52b}$$

$$\bar{Y} = [Y_{n,m} \,\vdots\, \underline{e}_{n}] \tag{52c}$$

From equation (27b) we know that

$$< \tilde{f}^{x}_{m,n} \,,\, S^{i}\underline{y}_{n} > = 0 \quad , \quad i = 1,2,\ldots,m \tag{53a}$$

$$< \tilde{f}^{x}_{m,n} \,,\, \underline{e}_{n} > = 0 \tag{53b}$$

Equation (53b) is satisfied only if we force $\tilde{f}^{x}_{m,n}(n) = 0$. This can always be done because the scalar ξ appears only in the last row of equation (51).

In particular, $\tilde{f}^{x}_{m,n}(n) = 0$ if we choose

$$\xi = -\sum_{i=1}^{m} a_{m}(i)\,x(n-i) \tag{54}$$

Since $\tilde{f}^{x}_{m,n}(n) = 0$, equation (53a) is seen to depend only on the first n-1 components of the vectors. It is easily seen, then, that $\tilde{f}^{x}_{m,n}(k)$ for k=1,2,...,n-1 are determined using the first n-1 rows of (51) in such a manner that the vector $[\tilde{f}^{x}_{m,n}(1),\ldots,\tilde{f}^{x}_{m,n}(n-1)]'$ is orthogonal to the first n-1 components of each column of $Y_{n,m}$. But this is exactly the problem of determining the forward prediction error based on n-1 data points. Thus, we see that

$$\tilde{f}^{x}_{m,n} = \begin{bmatrix} f^{x}_{m,n-1} \\ 0 \end{bmatrix} \tag{55}$$

Similar arguments show that this time annihilation property also holds for $\underline{f}^{y}_{m,n}$, $\underline{d}^{x}_{m,n}$, and $\underline{d}^{y}_{m,n}$ with resulting formulas similar to equation (55). We finally note that the scalars $\sigma_{m,n}$, $\tau_{m,n}$, $\mu_{m,n}$, and $\nu_{m,n}$ are formed as inner products of the prediction error vectors. Since the last element of $\tilde{f}^{x}_{m,n}$ (or $\tilde{f}^{y}_{m,n}$, $\tilde{d}^{x}_{m,n}$, $\tilde{d}^{y}_{m,n}$) is zero, it follows that

$$\sigma_{m,n-1} = [\tilde{f}^{x}_{m,n}]^{+}[\tilde{d}^{y}_{m,n}] \tag{56}$$

and similarly for $\tau_{m,n}$, $\mu_{m,n}$, and $\nu_{m,n}$.

With these thoughts in mind we are in a position to derive time update equations for σ, τ, μ, and ν. First, let us define the augmented matrices

$$\bar{X} = [X_{n,m} \,\vdots\, \underline{e}_{n}] \tag{57a}$$

$$\bar{Y} = [Y_{n,m} \,\vdots\, \underline{e}_{n}] \tag{57b}$$

Then it follows that

$$\sigma_{m,n-1} = [\tilde{f}^{x}_{m,n}]^{+}[\tilde{d}^{y}_{m,n}] = \underline{x}_{n}^{+}P^{c}_{\overline{XY}}(S^{m+1}\underline{y}_{n}) \tag{58}$$

where $\bar{X}$ and $\bar{Y}$ are defined by equation (57). Application of the projection operator theorem yields

$$\sigma_{m,n-1} = \underline{x}_{n}^{+}\left\{P^{c}_{XY} - P^{c}_{XY}\,\underline{e}_{n}[\underline{e}_{n}^{+}P^{c}_{XY}\underline{e}_{n}]^{-1}\underline{e}_{n}^{+}P^{c}_{XY}\right\}(S^{m+1}\underline{y}_{n})$$

$$\sigma_{m,n-1} = \sigma_{m,n} - f^{x}_{m,n}{}^{*}(n)[1-\gamma_{m,n}]^{-1}d^{y}_{m,n}(n) \tag{59}$$

where $1-\gamma_{m,n} \triangleq \underline{e}_{n}^{+}P^{c}_{XY}\,\underline{e}_{n}$ (60)

By rewriting equation (59) we arrive at the desired time update equation

$$\sigma_{m,n} = \sigma_{m,n-1} + \frac{[f^{x}_{m,n}(n)]^{*}[d^{y}_{m,n}(n)]}{1-\gamma_{m,n}} \tag{61}$$

The time update equation for τ is found in a similar manner by using $\bar{X}$ for A, $\bar{Y}$ for B, $S^{m+1}\underline{x}_n$ for $\underline{a}$, and $\underline{y}_n$ for $\underline{b}$ in equation (34) to yield

$$\tau_{m,n} = \tau_{m,n-1} + \frac{[d^{x}_{m,n}(n)]^{*}[f^{y}_{m,n}(n)]}{1-\gamma_{m,n}} \tag{62}$$

The update equations for μ and ν are found to be

$$\mu_{m,n} = \mu_{m,n-1} + \frac{[f^{x}_{m,n}(n)]^{*}[f^{y}_{m,n}(n)]}{1-\gamma_{m,n}} \tag{63}$$

$$\nu_{m,n} = \nu_{m,n-1} + \frac{[d^{x}_{m,n}(n)]^{*}[d^{y}_{m,n}(n)]}{1-\gamma_{m,n}} \tag{64}$$

$$f^x_{m+1,n}(n) = f^x_{m,n}(n) - b^x_{m,n-1}(n-1)\,\sigma^*_{m,n}/\omega^*_{m,n-1} \quad \text{(T-1)}$$

$$f^y_{m+1,n}(n) = f^y_{m,n}(n) - b^y_{m,n-1}(n-1)\,\tau_{m,n}/\omega_{m,n-1} \quad \text{(T-2)}$$

$$b^x_{m+1,n}(n) = b^x_{m,n-1}(n-1) - f^x_{m,n}(n)\,\tau^*_{m,n}/\mu^*_{m,n} \quad \text{(T-3)}$$

$$b^y_{m+1,n}(n) = b^y_{m,n-1}(n-1) - f^y_{m,n}(n)\,\sigma_{m,n}/\mu_{m,n} \quad \text{(T-4)}$$

$$\mu_{m+1,n} = \mu_{m,n} - \sigma_{m,n}\tau_{m,n}/\omega_{m,n-1} \quad \text{(T-5)}$$

$$\omega_{m+1,n} = \omega_{m,n-1} - \sigma_{m,n}\tau_{m,n}/\mu_{m,n} \quad \text{(T-6)}$$

$$\sigma_{m,n} = \sigma_{m,n-1} + [f^x_{m,n}(n)]^*[b^y_{m,n-1}(n-1)]/(1-\gamma_{m,n}) \quad \text{(T-7)}$$

$$\tau_{m,n} = \tau_{m,n-1} + [b^x_{m,n-1}(n-1)]^*[f^y_{m,n}(n)]/(1-\gamma_{m,n}) \quad \text{(T-8)}$$

$$\mu_{m,n} = \mu_{m,n-1} + [f^x_{m,n}(n)]^*[f^y_{m,n}(n)]/(1-\gamma_{m,n}) \quad \text{(T-9)}$$

$$\omega_{m,n} = \omega_{m,n-1} + [b^x_{m,n}(n)]^*[b^y_{m,n}(n)]/(1-\gamma_{m,n}) \quad \text{(T-10)}$$

$$\gamma_{m+1,n} = \gamma_{m,n} + [b^x_{m,n-1}(n-1)]^*[b^y_{m,n-1}(n-1)]/\omega^*_{m,n-1} \quad \text{(T-11)}$$

$$\gamma_{m+1,n+1} = \gamma_{m,n} + [f^x_{m,n}(n)]^*[f^y_{m,n}(n)]/\mu^*_{m,n} \quad \text{(T-12)}$$

Table 1: Summary of Update Equations

where p is the desired (maximum) autoregressive coefficient order.

Finally, when $m = 0$,

$$\gamma_{0,n} = 0 \quad (83)$$

Although other initial conditions may be obtained, these initial conditions are the only ones needed to implement the algorithm.

The implementation of the update formulas can be divided into three parts. First, for $n \le q$ the vector $\underline{y}_n$ is the zero vector, so no operations are performed. For $q+1 \le n \le q+p+1$, the maximum order m that can be used is $n-q-1$. In this time interval, as a new data point arrives not only are time updates performed, but also the model order is increased. For $n > q+p+1$, the model order remains at p and the time updates only are performed.

The implementation of the algorithm for $n \ge q+1$ is summarized below.

As the new data point becomes available:

1) Set $n \leftarrow n+1$

2) These quantities are available from the last iteration:

$f^x_{m,n-1}(n-1), f^y_{m,n-1}(n-1), b^x_{m,n-1}(n-1), b^y_{m,n-1}(n-1)$

$\sigma_{m,n-1}, \tau_{m,n-1}, \mu_{m,n-1}, \omega_{m,n-1}$

} for $m = 0, 1, \ldots, \min[p, n-q-2]$

3) New initial conditions:

$$f^x_{0,n}(n) = x(n)$$
$$b^x_{0,n}(n) = x(n)$$
$$f^y_{0,n}(n) = x(n-q)$$
$$b^y_{0,n}(n) = x(n-q)$$
$$\gamma_{0,n} = 0$$

If $n \le p+q+2$, $\quad \sigma_{n-q-1,n-1} = \tau_{n-q-1,n-1} = 0$

4) For each $m = 0,1,\ldots,\min[p-1, n-q-3]$ find

$\sigma_{m,n}, \tau_{m,n}, \mu_{m,n}, \omega_{m,n}$ using (T-7)-(T-10)

$f^x_{m+1,n}(n), f^y_{m+1,n}(n), b^y_{m+1,n}(n), b^y_{m+1,n}(n)$ using (T-1)-(T-4)

$\gamma_{m+1,n}$ using (T-11)

5) For $m = \min[p, n-q-2]$ find

$\sigma_{m,n}, \tau_{m,n}, \mu_{m,n}, \omega_{m,n}$ using (T-7)-(T-10)

6) If $n < p+q+1$ we need to add a filter order.

Set $m \leftarrow n-q-2$.

Find:

$f^x_{m+1,n}(n), f^y_{m+1,n}(n), b^x_{m+1,n}(n), b^y_{m+1,n}(n)$ using (T-1)-(T-4)

$\gamma_{m+1,n}$ using (T-1)

$\mu_{m+1,n}, \omega_{m+1,n}$ using (T-5), (T-6)

Set $m \leftarrow n-q-1$.

Find:

$\sigma_{m,n}, \tau_{m,n}$ using (T-7), (T-8) with $\sigma_{m,n-1} = \tau_{m,n-1} = 0$

At this point we are ready for the next data point to arrive.

It is clear from the above summary that $O(p)$ multiplications and additions are required to update the prediction errors. More specifically, in the time update mode [i.e., when $n > p+q+2$ so no filter orders need to be added] $14p$ multiplications and $10p$ additions are performed per update. In the time and order update mode, [i.e., when $n \le p+q+2$], $17p$ multiplications and $13p$ additions are performed. It should be noted that this computational requirement may be significantly reduced by using a normalized lattice form similar to that in [2].

IX. CONCLUSIONS

In this paper we have presented a recursive algorithm for obtaining the autoregressive coefficients of an ARMA model. The recursive algorithm is based on the prewindowed version of the high performance method of ARMA spectral estimation as described in Part 1. The recursive algorithm is computationally fast, requiring O(p) additions and multiplications to update the parameters. Moreover, the algorithm can be implemented using a lattice filter structure offering numerical robustness and nice convergence properties associated with lattice type algorithms.

We have not yet discussed the problem of recursively estimating the moving average coefficients in the ARMA model. We do not at this time have such an algorithm. However, it is worth noting that the moving average information is present in the output prediction error sequences, and the utilization of this information for moving average coefficient estimation is currently under study. Another area under study is the use of various normalization procedures to effect a decrease in computational requirements and in sensitivity.

Finally, we note that the recursive algorithm presented here is based on approximating a set of p Yule-Walker equations. It has been shown in Part 1 of this paper that for short data lengths, improved spectral estimates result from using more than p Yule-Walker equations. For those cases in which the amount of data is small, a fast recursive algorithm based on the approximation of t > p Yule-Walker equations would often prove useful. Such an algorithm is currently being pursued.

REFERENCES

[1] M. Morf and D. Lee, "Recursive Least Squares Ladder forms for Fast Parameter Tracking", Proc. 1978 IEEE Conf. D. & C., San Diego, CA, pp. 1362-1367.

[2] D. Lee and M. Morf, "Recursive Square Root Ladder Estimation Algorithms", Proc. 1980 IEEE Conf. Acoustics, Speech, and Signal Processing, Denver, CO, pp. 1005-1017.

[3] B. Friedlander, "Recursive Algorithms for Ladder Forms", SCI Technical Memorandum, 1980.

[4] B. Friedlander, "Recursive Algorithms for Pole-Zero Ladder Forms", SCI Technical Memorandum, 1980.

[5] J. Makhoul, "Stable and Efficient Lattice Methods for Linear Prediction, IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-25, No.5, October 1977, pp. 423-428.

[6] D. Lee, M. Morf, and B. Friedlander, "Recursive Least Squares Ladder Estimation Algorithms", IEEE Trans. Ckts. Sys., Vol. CAS-28, No.6, June 1981, pp. 467-481.

[7] B. Porat, B. Friedlander, and M. Morf, "Square Root Covariance Ladder Algorithms, Proc. 1981 IEEE Conf. Acoust. Speech, Sig. Proc., Atlanta, GA, pp. 877-880.

[8] M. Morf, B. Dickenson, T. Kailath, and A. Vieira, "Efficient Solution of Covariance Equations for Linear Prediction", IEEE Trans. Acoust. Speech, and Sig. Proc., Vol. ASSP-25, October 1977, pp. 429-433.

[9] K. Ogino, "Computationally Fast Algorithms for ARMA Spectral Estimation", Ph.D. Dissertation, Virginia Polytechnic Institute and State University, June 1981.

[10] E. Satorius and S. Alexander, "Channel Equalization Using Adaptive Lattice Algorithms", IEEE Trans. Comm, Vol. COMM-27, No. 6, June 1979, pp. 899-905.

[11] D. Falconer and L. Ljung, "Application of Fast Kalman Estimation to Adaptive Equalization, IEEE Trans. Comm. Vol. COMM-26, No. 10, October 1978, pp. 1439-1446.