OPTIMAL APPROXIMATE STOCHASTIC PARTIAL REALIZATION

Randolph L. Moses
Measurement and Control Group
Department of Electrical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands

We consider the problem of determining a stochastic realization from a given set of autocovariance samples. No structure (such as positive definiteness) is assumed on these given autocovariances. We present a method for determining a stable state space model whose output covariances minimize a weighted quadratic error function. The solution technique requires the use of an iterative minimization procedure, and a Gauss-Newton method is employed. Connections with Maxiumum Likelihood estimates are discussed, and a procedure for adaptively choosing an optimal quadratic error weight is described.

INTRODUCTION

This paper considers the following approximate stochastic partial realization (ASPR) problem: Given a sequence $\{R_k\}_0^M$ of autocovariance "measurements", and an integer n < M/2, find an n-th order state space model such that, when driven by white noise, its output covariances $\{r_k\}$ are "close to" R_k for $0 \leqslant k \leqslant M$ in some well-defined sense.

Specifically, we wish to find (A,B,C,D) in the n-th order state space model

$$x(k+1) = Ax(k) + Bw(k) \qquad\qquad (1.1)$$
$$y(k) = Cx(k) + Dw(k)$$

where $\{w(k)\}$ is white noise with zero mean and variance I. The function we wish to minimize is

$$F_Q = (\underline{R}_M - \underline{r}_M)^T Q (\underline{R}_M - \underline{r}_M) \qquad\qquad (1.2a)$$

where

$$\underline{R}_M = \begin{bmatrix} R_0 & R_1 & \cdots & R_M \end{bmatrix}^T \qquad\qquad (1.2b)$$

$$\underline{r}_M = \begin{bmatrix} r_0 & r_1 & \cdots & r_M \end{bmatrix}^T \qquad\qquad (1.2c)$$

and where Q is a non-negative definite weighting matrix.

It is important to note that the realized system is implicitly constrained to be asymptotically stable since the output sequence is assumed to be wide sense stationary. Also, note that no structure is assumed on the given $\{R_k\}$ samples. In particular, $\{R_k\}$ need not be non-negative definite, and need not fit a higher order linear model.

There are several important applications in which this ASPR problem arises. If $\{R_k\}$ is the covariance sequence associated with an ideal digital filter impulse response, then (A,B,C,D) realize an optimal n-th order filter design. In remote sensing and optical signal processing problems, one often wishes to obtain a spectral estimate or linear model based on noise-corrupted autocovariance measurements. Finally, in many time series analysis or stochastic system identification methods, one first estimates autocovariances from given data, then attempts to find a linear model which fits these covariances.

Several methods for obtaining approximate stochastic realizations have been proposed. However, many of these methods assume some structure on the given $\{R_k\}$ sequence, e.g. that it is exactly represented by a high order linear model or that it is part of a non-negative definite sequence [1-2]. If these restrictions are not met (as is often the case in practice), then such procedures generally do not guarantee stability of the realization. Also, they do not guarantee the existence of a solution to a Riccati equation (or, equivalently, they do not guarantee that an estimate of the power spectrum is factorable) [3-4]. Even if stability or factorability is ensured, no well-defined criterion is minimized. On the other hand, the solution to the optimal ASPR problem does guarantee stability and factorability while minimizing a quadratic error criterion; thus it can serve as a benchmark to which these suboptimal methods can be compared.

This paper first discusses the solution to the optimal ASPR problem for the case that Q in (1.2) is given. For simplicity we first consider the scalar output case. Since F_Q is a nonlinear function of the state space parameters, an iterative minimization procedure is employed. Sakai, et. al. [5] considers a similar problem, but restricts attention to an AR model. Gerdin [15] also treats a similar problem but uses a parameterization which does not guarantee spectral factorability.

Next, we consider the statistical properties of the ASPR estimate when the given autocovariance sequence is generated by a system of the form (1.1). We address the problem of adaptively determining Q in (1.2) so that minimization of F_Q yields asymptotically efficient estimates. This is accomplished by a three-step algorithm similar in concept to those in [8]. First a minimization of F for some given Q is performed. The realization obtained is used to determine an estimate of the optimum Q, and this new Q is used in a second minimization to obtain asymptotically efficient estimates. This procedure is applicable when the asymptotic covariance of the given covariance estimates are known functions of the state space parameters; in particular, it is applicable when the autocovariances are estimated from data. In the latter case this stochastic realization method can serve as an alternative to a maximum likelihood estimate that is obtained directly from time series data. For some problems, estimating autocovariances from data and then performing an optimal ASPR is a nearly efficient estimator that is less computationally burdensome than direct maximum likelihood methods. Finally, we discuss extensions to multivariable systems.

AN ITERATIVE SOLUTION PROCEDURE

In this section we develop a solution for the ASPR problem when Q is given. We first parametrize the state space quadruple. A Gauss-Newton minimization procedure employed, and equations for computing r_k and its partial derivatives are derived. An alternative parametrization using reflection coefficients is introduced to enable stability monitoring: in particular, asymptotic stability of the stochastic realization is ensured by simple bound constraints on the reflection coefficients.

Consider the state space model in (1.1) where w(k), $y(k) \in R$. Then the impulse

response of this system is

$$
h_k = \begin{cases} D & , k=0 \\ CA^{k-1}B & , k>0 \end{cases}
\qquad (2.1)
$$

The covariance of $\{y(k)\}$ is given by

$$
r_k \triangleq E\{y_{k+i}\, y_i\} = \sum_{i=0}^{\infty} h_i h_{i+k}
\qquad (2.2)
$$

Alternatively, we can express the covariance function in the frequency domain. Defining

$$
H(z) = \sum_{i=0}^{\infty} h_k z^{-k}
\qquad (2.3)
$$

$$
\Phi(z) = \sum_{i=-\infty}^{\infty} r_k z^{-k}
\qquad (2.4)
$$

it follows that

$$
\Phi(z) = H(z)\, H(z^{-1})
\qquad (2.5)
$$

In order to minimize F_Q it is necessary to express r_k as a function of a minimal set of parameters. The parametrization of the state space representation we will use is the observer form,

$$
A = \begin{bmatrix} 0 & 1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \\ \alpha_n & \cdots & & & \alpha_1 \end{bmatrix}
\qquad
B = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}
\qquad
C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}
\qquad (2.6)
$$

For this parametrization the transfer function $H(z)$ can be written as

$$
H(z) = \frac{b_0 + b_1 z^{-1} + \ldots + b_n z^{-n}}{1 + a_1 z^{-1} + \ldots + a_n z^{-n}} = \frac{B(z)}{A(z)}
\qquad (2.7a)
$$

where

$$
a_i = -\alpha_i \qquad i = 1,2,\ldots,n
\qquad (2.7b)
$$

$$
b_0 = D
\qquad (2.7c)
$$

$$
b_i = D\beta_i - \alpha_i \qquad i = 1,2,\ldots,n
\qquad (2.7d)
$$

From (2.7) it is clear that there is a 1-1 correspondence between the $\{a_i, b_i\}$ coefficients and the state-space parameters $\{\alpha_i, \beta_i, D\}$. It suffices to minimize F_Q as a function of $\underline{\theta}$ where

$$
\underline{\theta} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n & b_0 & b_1 & \cdots & b_n \end{bmatrix}^T
\qquad (2.8)
$$

Since F_Q is nonlinear in $\underline{\theta}$, an iterative minimization technique is employed. We will use the Gauss-Newton method; the k+1<u>st</u> minimization is then given by

$$\underline{\theta}^{k+1} = \underline{\theta}^k + (S^T Q S)^{-1} S^T Q \left[\underline{R}_M - \underline{r}_M(\underline{\theta}^k)\right] \qquad (2.9a)$$

where $\underline{R}_M$ and $\underline{r}_M$ are given in (1.2) and

$$[S]_{ij} = \left.\frac{\partial r_i}{\partial \theta_j}\right|_{\underline{\theta}^k} \qquad (m+1) \times (2n+1) \qquad (2.9b)$$

An initial θ^0 vector can be obtained from a standard realization method as in [4] [6], for example. This initial estimate may require modification to ensure stability.

In order to apply (2.9), we need only find r_k and S a function of $\underline{\theta}$. The representation of r_k for $0 \leqslant k \leqslant n$ is given by [7]

$$\underline{r}_n = A_1^{-1} B_1 A_2^{-1} \underline{b} \qquad (2.10a)$$

where

$$A_2 = \begin{bmatrix} 1 & & & & \\ a_1 & \ddots & & \bigcirc & \\ \vdots & \ddots & \ddots & & \\ a_n & \cdots & a_1 & 1 \end{bmatrix} \qquad (2.10b)$$

$$A_1 = A_2 + \begin{bmatrix} 0 & a_1 & \cdots & a_n \\ & \vdots & \ddots & \\ & a_n & \bigcirc & \\ 0 & & & 0 \end{bmatrix} \qquad (2.10c)$$

$$B_1 = \begin{bmatrix} b_0 & b_1 & \cdots & b_n \\ b_1 & & \ddots & \\ \vdots & \ddots & \bigcirc & \\ b_n & & & \end{bmatrix} \qquad (2.10d)$$

$$\underline{b} = \begin{bmatrix} b_0 & b_1 & \cdots & b_n \end{bmatrix}^T \qquad (2.10e)$$

For $k > n$, the covariances satisfy

$$r_k = -\sum_{i=1}^{n} a_i r_{k-i} \qquad (2.11)$$

Note that A_2^{-1} is lower triangular Toeplitz, and is given by

$$c \triangleq A_2^{-1} = \begin{bmatrix} c_0 & & & \\ c_1 & \ddots & \bigcirc & \\ \vdots & \ddots & \ddots & \\ c_n & \cdots & c_1 & c_0 \end{bmatrix} \qquad c_0 \equiv 1/b_0 \qquad (2.12a)$$

$$c_k = - \begin{bmatrix} c_{k-1} & \cdots & c_1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix} \qquad \text{.} \qquad (2.12b)$$

Formulas for the derivatives of r_k, $0 \leqslant k \leqslant n$, follow directly from (2.10) and the relation

$$\frac{\partial A^{-1}}{\partial x} = - A^{-1} \frac{\partial A}{\partial x} A^{-1} \qquad (2.13)$$

From (2.10a) and (2.13) we get

$$\frac{\partial \underline{r}_n}{\partial a_i} = - A_1^{-1} \frac{\partial A_1}{\partial a_i} \underline{r}_n + A_1^{-1} B_1 \frac{\partial C}{\partial a_i} \underline{b} \qquad (2.14)$$

The two derivative matrices are readily computed. It follows from (2.10c) that

$$\frac{\partial A_1}{\partial a_i} = \overbrace{\underbrace{\begin{bmatrix} 0 & & & \cdot & 1 \\ 1 & \cdot & & & \\ & \cdot & \cdot & & O \\ 0 & & \cdot & \cdot & \\ & & & \cdot & 1 \end{bmatrix}}_{n-i+1}}^{i+1} \qquad (2.15)$$

The derivative of C is Toeplitz and lower triangular, and can be recursively computed by

$$\frac{\partial c_k}{\partial a_i} = \begin{cases} 0 & , \quad k < i \\ -1 & , \quad k = i \\ c_{k-i} - \begin{bmatrix} \dfrac{\partial c_{k-1}}{\partial a_i} & \cdots & \dfrac{\partial c_i}{\partial a_i} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{k-i} \end{bmatrix} & , \quad k > i \end{cases} \qquad (2.16)$$

Also from (2.10a) we have

$$\frac{\partial \underline{r}_n}{\partial b_i} = A_1^{-1} \begin{bmatrix} h_i \\ \vdots \\ h_0 \\ \\ 0 \end{bmatrix} + \begin{bmatrix} b_i & \cdots & b_n \\ \vdots & \cdot & \\ b_n & & O \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_{n-i} \end{bmatrix} \qquad (2.17)$$

For $k > n$ the derivatives of r_k may be computed recursively. From (2.11) we have

$$\frac{\partial r_k}{\partial b_i} = - \sum_{j=1}^{n} a_j \frac{\partial r_{k-j}}{\partial b_i} \qquad , \quad k > n \qquad (2.18)$$

$$\frac{\partial r_k}{\partial a_i} = - r_{k-i} - \sum_{j=1}^{n} a_j \frac{\partial r_{k-j}}{\partial a_i} \qquad , \quad k > n \qquad (2.19)$$

is completely defined.

The above algorithm does not impose stability on the realization. Moreover, the minimum of F_Q is not unique in the b_i coefficients because reflecting a zero of $B(z)$ in or out of the unit circle does not affect F_Q. Both problems can be circumvented by reparameterizing F_Q in terms of reflection coefficients.

The reflection coefficients are defined from a vector

$$\underline{a}_n = \begin{bmatrix} a_{n,1} & a_{n,2} \cdots & a_{n,n} \end{bmatrix}^T$$

by the following recursion.

For $m=n,n-1,\ldots,1$:

$$\begin{cases} k_m = a_{m,m} \\ a_{m-1,i} = (a_{m,i} - k_m a_{m,m-i})/(1-k_m^2) \ , \ i=1,2,\ldots,m-1 \end{cases} \tag{2.20}$$

The inverse recursion is:

For $m=1,2,\ldots,n$

$$\begin{cases} a_{m,m} = k_m \\ a_{m,i} = a_{m-1,i} + k_m a_{m-1,m-i} \ , \ i=1,2,\ldots,m-1 \end{cases} \tag{2.21}$$

It is well-known that a polynomial $A(z)$ is stable (i.e. $A(z) = 0 \rightarrow |z| < 1$) if and only if $|k_i| < 1$ for $i=1,2,\ldots,n$. Thus, stability constraints are simple bound constraints using the reflection coefficients as parameters.

Define the new parameter vector as

$$\underline{\theta}' = \begin{bmatrix} k_1 & \cdots & k_n & b_0 & \cdots & b_n \end{bmatrix}^T \tag{2.22}$$

where $\{k_1 \cdots k_n\}$ are the reflection coefficients corresponding to $\{a_1 \cdots a_n\}$ in (2.8). (The b_i coefficients may be replaced by reflection coefficients in a similar manner if it is desired to ensure a minimum phase realization.

Clearly, $\underline{r}(\underline{\theta}')$ may be computed using (2.20) and (2.21). To compute S in terms of θ' we need the Jacobian matrix corresponding to the transformation (2.20). If we define

$$\begin{bmatrix} S' \end{bmatrix}_{ij} = \frac{\partial r_i}{\partial \theta'_j}$$

then

$$S' = S \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix} \tag{2.23}$$

where J is the Jacobian matrix

$$\begin{bmatrix} J \end{bmatrix}_{ij} = \frac{\partial a_i}{\partial k_j} \qquad (nxn)$$

It is readily found from (2.21) that

$$\frac{\partial a_i}{\partial k_j} = \begin{cases} 0 & , \; j > n \\ 1 & ; \; j = n \\ \dfrac{\partial a_{n-1,i}}{\partial k_j} + k_n \dfrac{\partial a_{n-1,n-i}}{\partial k_j} & , \; j < n \end{cases} \tag{2.24}$$

The above equations completely specify an iterative solution technique for the minimization of F_Q for fixed Q. This minimization produces a state-space realization of the form (1.2) (after transforming θ or θ' by use of (2.21) and (2.7)). Stability of the realization is assured by imposing bound constraints on the k_i coefficients. Moreover, the spectral factorization problem (or Riccati equation solution) is eliminated by parametrizing so that no spectral factorization need be computed (or no Ricatti equation need be solved).

MAXIMUM LIKELIHOOD ESTIMATES

In the previous section we derived a stochastic realization algorithm that uses a given, fixed weighting matrix Q. In this section we discuss the choice of Q and its relation to maximum likelihood (ML) estimates. Throughout this section we assume that the given autocovariances are obtained from a system of the form (1.1), i.e. that the given data is generated by a system in the model set.

Let us first assume that the given R_k sequence is a noisy sample of an n-th order state-space realization, where the noise is Gaussian with zero mean and known variance W, i.e.

$$\underline{R}_M - \underline{r}_M \sim N(\underline{0}, W) \tag{3.1}$$

Then the log-likelihood function for $\underline{r}_M$ is given by

$$L(\underline{r}_M) = -\frac{M+1}{2} \ln 2\pi - \frac{1}{2} \ln \det W - \frac{1}{2}(\underline{R}_M - \underline{r}_M)^T W^{-1} (\underline{R}_M - \underline{r}_M) \tag{3.2}$$

L is maximized by minimizing F_Q where $Q = W^{-1}$. Moreover, since there is a one-to-one correspondence between $\underline{r}$ and θ (assuming minimum phase B(z), for example), minimization of $F_{W^{-1}}$ corresponds to an ML-estimate of the parameters of the state-space realization given the "data" vector $\underline{R}_M$. Even if $\underline{R}_M$ is not Gaussian, minimization of $F_{W^{-1}}$ yields the minimum variance estimate for a large class of probability distributions.

In many cases of interest, however, the above assumptions are very restrictive. Generally, W is not known and it is not independent of the model parameters. Below we consider a less restrictive case, motivated by a similar problem in [8]. Assume that the R_k values are estimated from N data samples $y(1), \ldots, y(N)$. Assume further that

$$\sqrt{N} \; (\underline{R}_M - \underline{r}_M) \; \xrightarrow[N \to \infty]{\text{dist}} \; N(\underline{0}, W(\underline{\theta})) \tag{3.3}$$

where W is a known function of $\underline{\theta}$. Let $\hat{W}$ be a consistent estimate of $W(\underline{\theta})$, such that $\hat{W} - W$ is $O(1/\sqrt{N})$ (i.e. the covariance of $\hat{W} - W$ is asymptotically P/N). These assumptions on $\underline{R}_M$ and $\hat{W}$ are not restrictive; nearly all consistent estimates satisfy these order properties by the central limit theorem.

The normalized log-likelihood function for $\underline{r}_M$ is from (3.3)

$$\frac{1}{N} L(\underline{r}_M) = -\frac{M+1}{2N} \ln(2\pi) - \frac{1}{2N} \ln \det W(\underline{\theta}),$$

$$-\frac{1}{2} \left[R_M - \underline{r}_M(\theta)\right]^T W^{-1}(\theta) \left[(R_M - \underline{r}_M(\theta))\right] \tag{3.4}$$

Since $\underline{r}_M$ is a function of $\underline{\theta}$, (3.4) is minimized when

$$\frac{\partial \frac{1}{N} L(\underline{r}_M)}{\partial \underline{\theta}} = -\frac{1}{2N} \frac{\partial}{\partial \underline{\theta}} \left[\ln \det W\right] + S^T W^{-1}(R_M - \underline{r}_M)$$

$$-\frac{1}{2} \begin{bmatrix} (R_M - \underline{r}_M)^T \frac{\partial W^{-1}}{\partial \theta_1} & (R_M - \underline{r}_M) \\ \vdots \\ (R_M - \underline{r}_M)^T \frac{\partial W^{-1}}{\partial \theta_{2n+1}} & (R_M - \underline{r}_M) \end{bmatrix} = 0$$

Since $\hat{W} - W$ and $(R - \underline{r})$ are both $0(1/\sqrt{N})$, we have for fixed M

$$\frac{1}{N} \frac{\partial L(\theta)}{\partial \underline{\theta}} = S^T \hat{W}^{-1}\left(R_M - \underline{r}_M(\underline{\theta})\right) + 0(1/N) \tag{3.5}$$

Asymptotically (as $N \to \infty$) ML estimates are obtained by finding $\underline{\theta}$ such that $S^T \hat{W}^{-1}\left(R_M - \underline{r}_M(\underline{\theta})\right) = 0$. Such a $\underline{\theta}$ is found by minimizing $F_{\hat{W}^{-1}}$. Thus, an asymptotically efficient estimate $\underline{\theta}_{AML}$ for data satisfying (3.3) can be obtained by:

1. Obtaining a consistent estimate $\tilde{\theta}$ of $\underline{\theta}$ (using, e.g. a non-iterative technique or by minimizing F_Q for some fixed Q).

2. Computing $\hat{W} = W(\hat{\theta})$.

3. Minimizing $F_{\hat{W}^{-1}}$ using the Gauss-Newton algorithm to obtain $\hat{\underline{\theta}}_{AML}$.

Even though the repetition of steps 2 and 3 is unnecessary in the limit, repeating steps 2 and 3 for finite N may improve the variance of $\hat{\theta}_{AML}$. Also, for finite N the particular initial estimate of step 1 can affect the variance of $\hat{\theta}_{AML}$.

Note that the above procedure performs a minimization on M+1 "data" rather than the $y(1), \ldots, y(N)$ samples, which can significantly reduce computations if N >> M. Although $\underline{\theta}_{AML}$ is an asymptotically efficient estimate based on $\underline{R}_M$, $\underline{R}_M$ is not in general a sufficient statistic for $\{y(k)\}_1^N$. The obtained $\underline{\theta}_{AML}$ estimate may be "close enough" to an efficient estimate if the information lost in reducing from data samples to $\underline{R}_M$ is sufficiently small. One measure for the relative efficiency of ML estimates based on $\underline{R}_M$ and ML estimates based on $\{y(k)\}_1^N$ is the relative information index (RII) defined as [9]

$$RII(\underline{\theta}, M) = \frac{F(\underline{R}_M)}{F(Y)} \tag{3.6}$$

where $F(\underline{R}_M)$ and $F(Y)$ denote the Fischer information matrices corresponding to $\underline{R}_M$ and data, respectively. It has been shown that $0 \le RII \le 1$, and that RII = 1 if and only if $\underline{R}_M$ is a sufficient statistic for $\{y(k)\}_1^N$ [9]. If RII is "close to" 1 then efficient estimates based on $\underline{R}_M$ are "nearly" efficient estimates with respect to the data.

As before, we must choose a state space parametrization $\underline{\theta}$ such that:

   (1)   stability constraints can be (easily) imposed and

   (2)   the covariances can be expressed as an explicit function of $\underline{\theta}$.

Previously used paramterization [11-12] do not fulfill both of these requirements. A more appropriate parametrization for this problem is the form where

$$A = \text{diag}\left\{ \begin{bmatrix} 1 & 0 \\ -P_{10} & -P_{11} \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 \\ -P_{m_1} & -P_{m_0} \end{bmatrix} , p \right\} \quad , \quad m = \frac{n}{2} \tag{4.1}$$

The last element "p" of A appears only if n is odd. The matrices B and C are free with the exception that one element in each column of C (or row in B) is fixed, to one, say [14]. If the element fixed to one should be zero, then other elements in B or C will become large in magnitude during the minimization process, and the constraint can be changed to another element. The D matrix is lower triangular with positive diagonal elements. This representation is generically a feasible model. Note that the realization is stable if and only if each 2x2 block is stable (and p < 1 if n is odd). By parametrizing each block into two reflection coefficients, the stability constraints become bound constraints.

We also need to express $\underline{r}_M$ as a function of the state space parameter vector $\underline{\theta}$. Note that we can write

$$Y(z) = \left[C(zI-A)^{-1} B+D \right] U(z) \tag{4.2}$$

$$\Rightarrow \alpha(z) \, Y(z) = \left[C \, \text{adj}(zI-A)B + D\alpha(z) \right] U(z)$$

where $\alpha(z)$ is the characteristic polynomial of A. Since A is block diagonal, $\alpha(z)$ is the product of the characteristic polynomials of the blocks in A; moreover, $\text{adj}(zI-A)$ is easy to evaluate. Thus, from (4.2) we can find A_i and B_j in the relation

$$\sum_{i=0}^{n} A_i y_{k-i} = \sum_{j=0}^{n} B_j w_{k-j} \quad , \quad A_0 \equiv I \tag{4.3}$$

From (4.3) we can find an expression similar to (2.10). Following the derivation in [7], we get

$$\begin{bmatrix} A_0 & A_1 & \cdots & A_n \\ A_1 & & & \\ & & 0 & \\ A_n & & & \end{bmatrix} \begin{bmatrix} r_0 \\ \vdots \\ \vdots \\ r_n \end{bmatrix} + \begin{bmatrix} 0 & & & 0 \\ & A_0 & & \\ & \vdots & \ddots & \\ 0 & A_{n-1} & \cdots & A_0 \end{bmatrix} \begin{bmatrix} r_0^T \\ \vdots \\ \vdots \\ r_n^T \end{bmatrix} = \begin{bmatrix} D_0 \\ \vdots \\ \vdots \\ D_n \end{bmatrix} \tag{4.4}$$

where

$$D_k = \sum_{\ell=k}^{n} B_\ell H_{\ell-k}^T$$

$$H_k = \begin{cases} 0 & , k < 0 \\ \\ B_k + \sum_{i=1}^{n} A_i H_{k-i} & , k \geqslant 0 \end{cases}$$

Note that (4.4) can be solved in q steps by solving for the i-th columns of the r matrix and the r^T matrix in (4.4) at the i-th step. Also, from (4.2)-(4.4) we can compute the first derivatives of $\underline{r}_M$ with respect to $\underline{\theta}$ for use in a Gauss-Newton iteration.

The methods of section 3 can also be applied to the multivariable case. What is needed is an estimate of the optimal weighting matrix W and a relative information index for the specific problem at hand.

CONCLUSIONS

We have presented a method for obtaining stochastic realizations from a set of M+1 autocovariance "estimates". These estimates can come from data, direct (noisy) measurement, or high order models; however, no structure is assumed on them. Moreover, the realization obtained is one which minimizes a weighted quadratic error criterion and which is guaranteed to be stable. The algorithm incorporates an iterative Gauss-Newton minimization procedure.

Also discussed was the use of this algorithm in obtaining maximum likelihood and asymptotically efficient estimates. Specifically, when a large number of data samples are available, one can compute sample covariances and then obtain a stochastic realization from them instead of from data. The advantage of such a scheme is that it may be computationally faster than direct maximum likelihood methods. Moreover, the information lost in reducing the data to sample covariances can be measured and controlled, so one can obtain realizations whose variances are asymptotically as close to efficient as desired.

REFERENCES

[1]   U.B. Desai and D. Pal, A Realization Approach to Stochastic Model Reduction Using Balanced Stochastic Realizations, 21st IEEE Conf. Decision and Control, (1982) Orlando, FL, 1105-1112.

[2]   G. Menga and N. Sundararajan, Stochastic Modeling of Mean-Wind Profiles for In-Flight Estimation - A New Approach to Lower Order Stochastic Realization Schemes, IEEE Trans. Automatic Control, AC-26, 3 (1976) 226-38.

[3]   R.L. Moses, Design and Analysis of Fast Recursive ARMA Spectral Estimators, Ph.D. Diss., Virginia Polytechnic Institute and State University (June 1984).

[4]   Proc. 2nd ASSP Workshop on Spectral Estimation (1983) Tampa, FL.

[5]   H. Sakai, K. Orita and N. Iwama, AR Spectrum Analysis Based on Noisy Autocovariance Sequence, Proc. Int. Conf. Acoustics, Speech and Signal Processing, (1982) Paris, 1030-1033.

[6]   K.S. Arun and S.Y. Kung, State Space Modeling and Approximate Realization Methods for ARMA Spectral Estimation, IEEE Int. Conf. Systems, Man and Cybernetics, (1984) Bombay/Delhi, India.

[7]   J.P. Dugré, A. Beex, and L. Scharf, Generating Covariance Sequences and the Calcualtion of Quantization and Roundoff Error Variances in Digital Filters, IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-28, 1, (1980) 103-104.

[8]   T. Söderström and P. Stoica, Instrumental Variable Methods for System
        Identification, (Springer Verlag, 1983).

[9]   S. Bruzzone and M. Kaveh, A Criterion for Selecting Information-Preserving
        Data Reductions for Use in the Design of Multiple Parameter Estimators,
        IEEE Trans. Information Theory, IT-29, 3 (1981) 466-70.

[10]   S. Bruzzone and M. Kaveh, Information Tradeoffs in Using the Sample Auto-
        correlation Function in ARMA Parameter Estimation, IEEE Trans. Acoustics,
        Speech, and Signal Processing, ASSP-32, 4 (1984), 701-715.

[11]   E. Tse and H. Weinert, Structure Determination and Parameter Estimation
        for Multivariable Stochastic Linear Systems, IEEE Trans. Automatic Con-
        trol, AC-25, 5 (1975) 603-613.

[12]   C. Kubrusly and A. Lucena, A Comparison of Three Covariance Techniques for
        System Identification, Int. J. Sys. Sci., 14, 1 (1983) 31-51.

[13]   J. Monahan, Reparametrizations for the Stability Region in ARMA Time
        Series Models, Research Mem. No. 244, The Inst. Stat. Math., N.C. State
        University (Oct. 1982).

[14]   L. Keviczky, J. Bokor and S. Veres, Strong Consistency of ML Estimators
        using Partial Fraction and Elementary Subsystem Representation of Multi-
        variable Systems, IFAC 9th World Congress (1984) Budapest.

[15]   K. Gerdin, Equivalent Filter Identification from Autocorrelation Data,
        Proc. of the 3rd IFAC Symposium on Identification and Parameter Estimation
        (1973) The Hague/Delft.