The Annals of Statistics
1978, Vol. 6, No. 3, 643-648

COVARIANCE CHARACTERIZATION BY PARTIAL AUTOCORRELATION MATRICES[1]

BY M. MORF, A. VIEIRA AND T. KAILATH

Stanford University

It is known that the autocorrelation function of a stationary discrete-time scalar process can be uniquely characterized by the so-called partial autocorrelation function, which is a sequence of numbers less or equal to one in magnitude. We show here that the matrix covariance function of a multivariate stationary process can be characterized by a sequence of matrix partial correlations, having *singular values* less than or equal to one in magnitude. This characterization can be used to extend to the multivariate case the so-called maximum entropy spectral analysis method.

1. Introduction. It has been known for some time (see, e.g., [1, 2, 10]) that the autocorrelation function (ACF) for a scalar stationary discrete-time process also has a useful characterization in terms of the so-called partial autocorrelation coefficients (also called in [10] the partial autocorrelation function or PACF). An important application of this characterization has been to the development of a new spectral estimation technique known as the (Burg) maximum entropy method, in which the PACF (rather than the ACF) is estimated by minimizing the sum of the squares of the forward and backward prediction errors (or innovations).

No completely satisfactory vector (or multivariable) extension of these results seems to be known as yet, though some results have been reported by Burg [2], Jones [5], and Nuttall [8]. In [7] we have taken a somewhat different route by first finding the vector multivariate analog of the ACF—PACF correspondence, which then enables a natural extension of the above-mentioned Burg technique. Since this extension is perhaps of more interest in applications, we shall treat it elsewhere. In this note, we present only the vector generalization of the ACF-PACF correspondence, which may have theoretical implications beyond its role in suggesting spectral-estimation methods.

In the scalar case, the PACF is naturally encountered in the efficient solutions of Levinson [6] and Durbin [3] to the problem of fitting autoregressive (AR) models to a given stationary autocorrelation sequence. In this method, AR models of increasing orders are fitted by calculating the one-step forward and backward predictors at each stage as a linear combination of the corresponding

Received December 1976; revised August 1977.

[1] This work was supported in part by the Defense Communications Agency, Contract DCA100-76-R-0136, by the National Science Foundation under Contract NSF Eng-75-18962 and the Air Force Office of Scientific Research, AF Systems Command, under Contract AF44-620-74-C-0068.

AMS 1970 *subject classifications.* Primary 62M10, 62N15, 62M15, 60G10.

Key words and phrases. Partial autocorrelation matrices, multivariate stationary processes, multivariate maximum entropy method of spectral analysis.

quantities at the previous stage. The coefficients in this linear combination are particularly important and give a set of partial autocorrelation coefficients (also called reflection coefficients) which turn out to have a one-to-one correspondence to the given autocorrelation coefficients. Now multivariate extensions of the Levinson–Durbin algorithm have been obtained by Whittle [11] and Wiggins and Robinson [12]. We shall show that a normalized version of this algorithm, which we shall call the LWR algorithm, provides the vector generalization of the previous characterization theorem: the matrix autocorrelation coefficients of a vector stationary process have a one-to-one correspondence to a sequence of (partial autocorrelation) matrices whose singular values have magnitude less than or equal to unity. (We recall that the singular values of a matrix A are the positive square roots of the eigenvalues of AA'.)

One application of this result is to the problem of suitably extending a given correlation sequence $\{R_0, \cdots, R_n\}$ so that the new sequence is also a correlation sequence (i.e., has the right positive-definiteness properties). It is not easy to characterize all suitable extensions $\{R_{n+1}, R_{n+2}, \cdots\}$ but, with the above result, we know that in the PACF domain we just have to add any matrix with singular values less than unity. Applications to spectral estimation have already been briefly mentioned (see also Section 3).

2. The normalized LWR algorithm. Suppose we are given the $(N+1)$ matrices[2] of size $m \times m$

$$R_n = E\{y_{t+n} y_t'\}, \quad |n| \leqq N$$

where $\{y_t\}$ is an m-vector stationary random process, so that we also have

$$R_{-n} = R_n' .$$

Then the so-called autoregressive (or maximum entropy) extension of the sequence $\{R_n, |n| \leqq N\}$ is known (Parzen [9], Burg [2]) to be defined by the expressions

$$\begin{aligned} R(z) \equiv \textstyle\sum_{n=-\infty}^{\infty} R_n z^{-n} &= A_N^{-1}(z) R_N^{\varepsilon} A_N^{-T}(z^{-1}) \\ &= B_N^{-1}(z) R_N^{r} B_N^{-T}(z^{-1}) \end{aligned}$$

where $A_N(z)$ and $B_N(z)$ are the forward and backward prediction filters, and R_N^{ε} and R_N^{r} the respective prediction error (or innovation) variances, defined by the equations (see, e.g., [11])

$$\begin{bmatrix} I, A_{N,1}, \cdots, A_{N,N} \\ B_{N,N}, \cdots, B_{N,1}, I \end{bmatrix} \mathbb{R}_N = \begin{bmatrix} R_N^{\varepsilon}, 0, \cdots, 0 \\ 0, \cdots, 0, R_N^{r} \end{bmatrix} \tag{1}$$

where

$$\mathbb{R}_N \equiv \begin{bmatrix} R_0 & R_1 & \cdots R_N \\ R_{-1} & R_0 & \cdots R_{N-1} \\ \vdots & & \\ R_{-N} & R_{-N+1} \cdots & R_0 \end{bmatrix}$$

[2] A' denotes the transpose of A. If A is complex then it will denote conjugate complex of A. Also $A^{-T} \equiv (A')^{-1}$.

and $\{A_{N,i}\}$ and $\{B_{N,i}\}$ are the coefficients in the forward and backwards prediction filters, respectively, i.e.,

$$A_N(z) \equiv I + A_{N,1}z^{-1} + \cdots + A_{N,N}z^{-N} \tag{2a}$$

$$B_N(z) \equiv B_{N,N} + B_{N,N-1}z^{-1} + \cdots + Iz^{-N}. \tag{2b}$$

Note that (1) are just the (Yule–Walker) equations obtained by minimizing trace $E\{\varepsilon_{N,t}\ \varepsilon'_{N,t}\}$ and trace $E\{r_{N,t}\ r'_{N,t}\}$, where

$$\varepsilon_{N,t} \equiv y_t + A_{N,1}y_{t-1} + \cdots + A_{N,N}y_{t-N} \tag{3a}$$

$$r_{N,t} \equiv B_{N,N}y_t + \cdots + B_{N,1}y_{t-N+1} + y_{t-N} \tag{3b}$$

are respectively the forward and backward prediction errors.

LEMMA 1. *We have*

$$\det R_n = \prod_{i=0}^{n} \det R_i^{\varepsilon} = \prod_{i=0}^{n} \det R_i^{r}, \quad |n| \leqq N$$

so that

$$\det R_n^{\varepsilon} = \det R_n^{r}, \quad |n| \leqq N$$

and

$$R_n > 0 \Leftrightarrow R_i^{\varepsilon} > 0 \Leftrightarrow R_i^{r} > 0, \quad i = 0, 1, 2, \cdots, n.$$

PROOF. These relations follow from the readily verified identity

$$\left[\begin{array}{c|c} I & A_{1,n} \cdots A_{n,n} \\ \hline 0 & I \end{array}\right] \mathbb{R}_n \left[\begin{array}{c|c} I & 0 \\ \hline A'_{1,n} & \\ \vdots & I \\ A'_{n,n} & \end{array}\right] = \left[\begin{array}{c|c} R_n^{\varepsilon} & 0 \\ \hline 0 & R_{n-1} \end{array}\right]$$

and a similar one with B_n and R_n^{r}. □

The LWR *algorithm.* Equations (1) can be solved in an efficient recursive manner by using the LWR algorithm (see, e.g., [11], [12]),

$$\begin{aligned} A_{n+1} &\equiv [I, A_{n+1,1}, \cdots, A_{n+1,n+1}] \\ &= [I, A_{n,1}, \cdots, A_{n,n}, 0] - \Delta_{n+1}(R_n^{r})^{-1}[0, B_{n,n}, \cdots, B_{n,1}, I], \end{aligned} \tag{7}$$

$$\begin{aligned} B_{n+1} &\equiv [B_{n+1,n+1}, \cdots, B_{n+1,1}, I] \\ &= [0, B_{n,n}, \cdots, B_{n,1}, I] - \Delta'_{n+1}(R_n^{\varepsilon})^{-1}[I, A_{n,1}, \cdots, A_{n,n}, 0], \end{aligned} \tag{8}$$

where

$$\begin{aligned} \Delta_{n+1} &\equiv R_{n+1} + A_{n,1}R_n + \cdots + A_{n,n}R_1 \\ &= (R_{-n-1} + B_{n,1}R_{-n} + \cdots + B_{n,n}R_{-1}), \end{aligned} \tag{9}$$

and

$$R_{n+1}^{\varepsilon} = R_n^{\varepsilon} - \Delta_{n+1}(R_n^{r})^{-1}\Delta'_{n+1}, \tag{10}$$

$$R_{n+1}^{r} = R_n^{r} - \Delta'_{n+1}(R_n^{\varepsilon})^{-1}\Delta_{n+1} \tag{11}$$

$$A_0 = B_0 = I, \qquad R_0^{\varepsilon} = R_0^{r} = R_0.$$

The normalized LWR *algorithm.* It will be useful to introduce a normalized form of this LWR algorithm. For this we shall need various matrix square roots: for any positive-definite matrix R, define a lower triangular matrix[3] $R^{\frac{1}{2}}$ such that

$$R^{\frac{1}{2}}R^{T/2} = R\,. \tag{12}$$

Here $A^{T/2} = (A^{\frac{1}{2}})'$; also let $A^{-\frac{1}{2}} = (A^{\frac{1}{2}})^{-1}$, $A^{-T/2} = (A^{-\frac{1}{2}})$. $R^{\frac{1}{2}}$ can be made unique by requiring the diagonal elements to be positive. Recalling (12) let us now define

$$P_{n+1} \equiv (R_n^{\varepsilon})^{-\frac{1}{2}}(R_{n+1}^{\varepsilon})^{\frac{1}{2}}\,; \qquad Q_{n+1}^{\frac{1}{2}} \equiv (R_n^{r})^{-\frac{1}{2}}(R_{n+1}^{r})^{\frac{1}{2}} \tag{13}$$
$$P_0^{\frac{1}{2}} = Q_0^{\frac{1}{2}} = R_0^{\frac{1}{2}}$$

and

$$\rho_{n+1} \equiv (R_n^{\varepsilon})^{-\frac{1}{2}}\Delta_{n+1}(R_n^{r})^{-T/2}\,. \tag{14}$$

Then from (10)—(11) we can see with a little calculation that

$$P_{n+1} = I - \rho_{n+1}\rho'_{n+1} \tag{15}$$

and

$$Q_{n+1} = I - \rho'_{n+1}\rho_{n+1}\,. \tag{16}$$

The recursions (7)—(8) can then be written in the normalized form

$$\tilde{A}_{n+1}(z) = P_{n+1}^{-\frac{1}{2}}[\tilde{A}_n(z) - z^{-1}\rho_{n+1}\tilde{B}_n(z)]\,, \qquad \tilde{A}_0(z) = R_0^{-\frac{1}{2}} \tag{17}$$

$$\tilde{B}_{n+1}(z) = Q_{n+1}^{-\frac{1}{2}}[z^{-1}\tilde{B}_n(z) - \rho'_{n+1}\tilde{A}_n(z)]\,, \qquad \tilde{B}_0(z) = R_0^{-\frac{1}{2}} \tag{18}$$

where

$$\tilde{A}_n(z) \equiv (R_n^{\varepsilon})^{-\frac{1}{2}}A_n(z)\,, \qquad \tilde{B}_n(z) \equiv (R_n^{r})^{-\frac{1}{2}}B_n(z)\,.$$

To complete the normalized algorithm, define

$$\tilde{\Delta}'_{n+1} = \tilde{B}_{n,0}R_{-n-1} + \tilde{B}_{n,1}R_{-n} + \cdots + \tilde{B}_{n,n}R_{-1} = (R_n^{r})^{-\frac{1}{2}}\Delta'_{n+1}\,. \tag{19}$$

Then we see that

$$\rho_{n+1} = P_n^{-\frac{1}{2}} \cdots P_0^{-\frac{1}{2}}\tilde{\Delta}_{n+1}\,. \tag{20}$$

For, using (13), we can write

$$(R_n^{\varepsilon})^{\frac{1}{2}} = P_0^{\frac{1}{2}} \cdots P_n^{\frac{1}{2}}\,, \qquad (R_n^{r})^{\frac{1}{2}} = Q_0^{\frac{1}{2}} \cdots Q_n^{\frac{1}{2}}$$

which, with (14), leads to the formula (20).

Equations (15)—(20) then form a complete set of recursions that are readily seen to be equivalent to the recursions (7)—(11): $A_n(z)$ and $B_n(z)$ can be recovered from $\tilde{A}_n(z)$ and $\tilde{B}_n(z)$ by

$$A_n(z) = (\tilde{A}_{n,0})^{-1}\tilde{A}_n(z) \tag{21}$$

$$B_n(z) = (\tilde{B}_{n,0})^{-1}\tilde{B}_n(z) \tag{22}$$

and R_n^{ε}, R_n^{r} from (21).

[3] Even though a square root may be in general nontriangular we require such a form so as to get a uniquely defined matrix.

The significance of this equivalent form of the LWR algorithm is that we can now uniquely characterize the covariance $\{R_n, |n| \leqq N\}$ by the partial correlation sequence $\{\rho_n, n = 1, \cdots, N\}$ and R_0. In other words we have the following theorem.

THEOREM 1. *The normalized* LWR *establishes a one-to-one correspondence between* $\{\mathbb{R}_n, |n| \leqq N\}$ *with* $\mathbb{R}_n > 0$ *and* $\{R_0; \rho_n, n = 1, \cdots, N\}$, *where* $\{\rho_n, n = 1, \cdots, N\}$ *is a set of matrices with singular values* $\{\sigma_i(\rho_n)\}_{i=1,\cdots,m}$ *less than one in magnitude.*

PROOF. Given $\mathbb{R}_n > 0$, the partial correlation sequence ρ_n, $n = 1, \cdots, N$ is uniquely defined by (14). Also from Lemma 1, $R_n^e > 0$, $n = 1, \cdots, N$. But from (15), $R_n^e > 0 \Leftrightarrow |\sigma_i(p_n)| < 1$, $i = 1, \cdots, m$. Conversely given $\{R_0; \rho_n, n = 1, \cdots, N\}$ define $\{P_n^{\frac{1}{2}}, Q_n^{\frac{1}{2}}, n = 1, \cdots, N\}$ through (15)—(16). These square roots always exist if $\{|\sigma_i(\rho_n)| < 1, i = 1, \cdots, m\}$. Then, Equations (17)—(18) uniquely define $\{\tilde{A}_n(z), \tilde{B}_n(z), n = 1, \cdots, N\}$, which in turn uniquely define $\mathbb{R} > 0$ from the LWR algorithm. □

NOTE. If for some n the covariance matrix R_n is not strictly positive (i.e., $|R_n| = 0$), the LWR recursion will stop. In this case ρ_n will have singular values equal to one in magnitude.

3. Concluding remarks. Theorem 1 gives a parametrization of the correlation function of a multivariate stationary process in terms of matrices with singular values less than one in magnitude. This generalizes a result known for the covariance function of a scalar stationary process.

We may remark here that the scalar case has also been considered in the context of orthogonal polynomial theory [4] and consequently the result given here carries over to matrix orthogonal polynomials.

The implication for spectral estimation is that instead of estimating $\{R_n\}$ directly we can first estimate $\{\rho_n\}$ and then find the corresponding $A_n(z)$ (and if desired the corresponding R_n). Finding $\{R_n\}$ from $\{\rho_n\}$ can be done efficiently through formulas (19)—(20). In [7] we give estimates of $\{\rho_n\}$ that are guaranteed to have singular values in magnitude less than or equal to one, so that the corresponding R_n form a nonnegative-definite sequence.

REFERENCES

[1] BARNDORFF-NIELSEN, O. and SCHOU, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *J. Multivariate Anal.* **3** 408-419.

[2] BURG, J. P. (1975). Maximum entropy spectral analysis. Ph. D. dissertation, Geophysics Dept., Stanford Univ.

[3] DURBIN, J. (1960). The fitting of time-series models. *Rev. Inst. Internat. Statist.* **28** 233-244.

[4] GERONIMUS, YA. L. (1960). *Polynomials Orthogonal on a Circle and Interval.* Pergamon Press, Oxford.

[5] JONES, R. H. (1976). Multivariate maximum entropy spectral analysis. *Applied Time Series Analysis Symposium* (Tulsa, Oklahoma).

[6] LEVINSON, N. (1947). The Wiener RMS (root-mean-square) error criterion in filter design and prediction. *J. Math. Phys.* **25** 261-278.

[7] MORF, M., VIEIRA, A., LEE, D. and KAILATH, T. (1977). Recursive multichannel maximum entropy method. *Proc. JACC* 113–117 (San Francisco).

[8] NUTTALL, A. H. (1976). Multivariate linear predictive spectral analysis employing weighted forward and backward averaging: a generalization of Burg's algorithm. Naval Underwater Systems Center Technical Report No. 5501.

[9] PARZEN, E. (1968). Multiple time-series modeling. *Proc. Second Internat. Symp. Multivariate Anal.* (Dayton, Ohio).

[10] RAMSEY, F. (1974). Characterization of the partial autocorrelation function. *Ann. Statist.* **2** 1296–1301.

[11] WHITTLE, P. (1963). On the fitting of multivariable autoregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50** 129–134.

[12] WIGGINS, R. A. and ROBINSON, E. A. (1965). Recursive solution to the multichannel filtering problem. *J. Geophys. Res.* **70** 1885–1891.

INFORMATION SYSTEMS LABORATORY
DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305