

Multiwindow Estimators of Correlation

L. Todd McWhorter and Louis L. Scharf, *Fellow, IEEE*

Abstract—This paper is concerned with the structure of estimators of correlation matrices and correlation sequences. We argue that reasonable estimators of the correlation matrix are quadratic in the data and nonnegative definite. We also specify the structure of the estimator when the data are modulated: a property we call modulation covariance. We state a representation theorem for estimators that have these attributes. We also derive a representation for estimators that have the additional requirement that the estimated matrix be Toeplitz. These representation theorems admit estimators that use multiwindowed copies of the data. In many circumstances, this multiwindow structure is superior to the conventional sum of lagged-products or outer-product estimators.

I. INTRODUCTION

IT IS commonplace for adaptive detectors, estimators, and filters to have embedded within them estimators of correlation. In time series and array processing problems, the estimators are usually sums of outer products of measurement vectors or sums of lagged inner products. In linear prediction, they are Gramians associated with the various methods of linear prediction.

In this paper, we pose the question: “What is the most general estimator of correlation one can consider, under the constraints that it be quadratic in the data, modulation covariant, Hermitian symmetric, and perhaps Toeplitz?” We make the case for each of these constraints and then state a representation theorem that says every such estimator of correlation admits a multiwindow representation. We show that commonly used estimators are special cases of multiwindow estimators and suggest that each of them may be improved with multiwindowing. We derive a general mean-squared error bound for multiwindow estimators that shows how a multiplicity of windows may be used to improve the bound. We conclude with a numerical example.

The multiwindow idea first arose in spectrum analysis in a paper by Thomson [17], where data were windowed in order to reduce variance in exchange for poorer resolution (larger bias). Multiwindow estimators of correlation produce similar effects [9], [13]. In fact, when coupled with Blackman-Tukey spectral estimators [2], they reproduce Thomson’s multiwindow spectrum estimators. We assert that the bias-variance trade is fundamental in signal processing and speculate that the

applications for the multiple window idea abound in parameter identification, moment estimation, and other areas that rely on estimates of correlation.

II. PROPERTIES OF THE CORRELATION MATRIX

A reasonable strategy for designing any estimator is to require that the estimator possess certain desirable attributes, find the largest class of estimators that have the required attributes, and then optimize within this class. The objective of this section is to apply this strategy to the design of estimators of the correlation matrix. A few crucial properties of the true correlation matrix will provide the guidelines for the design of our estimators.

Let $\mathbf{y} = [y_0 \ y_1 \ \cdots \ y_{N-1}]^T$ denote a vector of complex-valued wide-sense stationary (WSS) data with power spectrum $S(e^{j\theta})$. The $N \times N$ correlation matrix associated with the data is

$$\begin{aligned} \mathbf{R} &= E\{\mathbf{y}\mathbf{y}^*\} = \int_{-\pi}^{\pi} S(e^{j\theta}) \boldsymbol{\psi}(e^{j\theta}) \boldsymbol{\psi}^*(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= \begin{bmatrix} r_0 & r_{-1} & \cdots & r_{1-N} \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{N-1} & \cdots & r_1 & r_0 \end{bmatrix} \end{aligned} \quad (1)$$

where $\boldsymbol{\psi}(e^{j\theta}) = [1 \ e^{j\theta} \ \cdots \ e^{j(N-1)\theta}]^T$ is a DTFT vector of length N . The power in the time series is

$$r_0 = \frac{1}{N} \text{tr}\{\mathbf{R}\} = \int_{-\pi}^{\pi} S(e^{j\theta}) \frac{d\theta}{2\pi}.$$

This power may be “resolved” with quadratic forms like $\mathbf{a}^* \mathbf{R} \mathbf{a}$

$$\begin{aligned} \mathbf{a}^* \mathbf{R} \mathbf{a} &= E\{(\mathbf{a}^* \mathbf{y})(\mathbf{a}^* \mathbf{y})^*\} = \int_{-\pi}^{\pi} S(e^{j\theta}) |A(e^{j\theta})|^2 \frac{d\theta}{2\pi} \geq 0 \\ A(e^{j\theta}) &= \boldsymbol{\psi}^*(e^{j\theta}) \mathbf{a}: \text{complex frequency response of filter } \mathbf{a}. \end{aligned}$$

Since the power spectrum $S(e^{j\theta})$ is nonnegative, the corresponding correlation matrix is nonnegative definite.

Let $\mathbf{D} = \text{diag}\{1, e^{j\phi}, \dots, e^{j(N-1)\phi}\}$ denote an $N \times N$ modulation matrix with modulation frequency ϕ . The correlation matrix \mathbf{R} has the following properties:

- P1** quadratic scaling: $E\{(\alpha \mathbf{y})(\alpha \mathbf{y})^*\} = |\alpha|^2 \mathbf{R}$;
- P2** nonnegativity: $\mathbf{R} \geq \mathbf{0}$;
- P3** modulation covariance: $E\{(\mathbf{D}(e^{j\phi}) \mathbf{y})(\mathbf{D}(e^{j\phi}) \mathbf{y})^*\} = \mathbf{D}(e^{j\phi}) \mathbf{R} \mathbf{D}(e^{-j\phi})$.

Manuscript received January 13, 1996; revised September 3, 1997. This work was supported by the Office of Naval Research, Statistics and Probability Branch, under Contract N00014-89-J1070. The associate editor coordinating the review of this paper and approving it for publication was Dr. Monique Fargues.

L. T. McWhorter was with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309 USA. He is now with Orincon Corporation, San Diego, CA 92121 USA.

L. L. Scharf is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309 USA.

Publisher Item Identifier S 1053-587X(98)01336-1.

We submit that any practical *estimator* of the correlation matrix must also possess properties P1)–P3). It seems reasonable to require an estimator of the correlation matrix to be nonnegative definite and to have the quadratic scaling property. However, it may not be immediately obvious why the modulation covariance property is important. In the following, we demonstrate that ESPRIT implicitly relies on estimators of the correlation matrix that have the modulation covariance property. We also show that modulation covariance is equivalent to requiring the estimator to use only “proper” lag-products of the data.

A. Modulation Covariance and ESPRIT

In a number of estimation problems, including deterministic modal analysis and DOA estimation, a snapshot of observed data $\mathbf{y}_k \in \mathbb{C}^N$ can be modeled as a sum of sinusoidal signals and noise [16]

$$\mathbf{y}_k = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_k + \mathbf{n}_k \quad k = 0, 1, \dots, M-1.$$

The signal component is assumed to be a vector in a low-dimensional subspace defined by

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}) &= [\boldsymbol{\psi}(e^{j\theta_1}) \ \boldsymbol{\psi}(e^{j\theta_2}) \ \dots \ \boldsymbol{\psi}(e^{j\theta_p})] \in \mathbb{C}^{N \times p} \\ &= [\boldsymbol{\psi}(z_1) \ \boldsymbol{\psi}(z_2) \ \dots \ \boldsymbol{\psi}(z_p)]; \quad z_k = e^{j\theta_k}. \end{aligned}$$

That is, the signal subspace is defined by the deterministic Vandermonde matrix $\mathbf{A}(\boldsymbol{\theta})$.

There exist a number of algorithms for estimating either $\{z_i\}$ or $\{\theta_i\}$ given the snapshots $\{\mathbf{y}_k\}$. As a representative algorithm, we choose ESPRIT. We consider an estimator or algorithm to be modulation covariant if when we modulate the data, the estimators then produce shifted or modulated versions of the estimates generated using the original, unmodulated, data. In the following, we demonstrate that if $\hat{\mathbf{R}}(\mathbf{y})$ is modulation covariant, then ESPRIT produces appropriately modulated estimates.

Begin by partitioning the estimated correlation matrix as

$$\hat{\mathbf{R}}(\mathbf{y}) = \begin{bmatrix} \hat{r}_0 & \hat{\mathbf{r}}_1^* \\ \hat{\mathbf{r}}_1 & \hat{\mathbf{H}}(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{r}}_0^* & \hat{r}_{N-1} \\ \hat{\mathbf{G}}(\mathbf{y}) & \hat{\mathbf{r}}_N \end{bmatrix}.$$

At its simplest, the ESPRIT algorithm [12] estimates the modes $\{z_i\}$ from the generalized eigenvalues of $\hat{\mathbf{G}}(\mathbf{y})$ and $\hat{\mathbf{H}}(\mathbf{y})$. That is

$$[\hat{\mathbf{G}}(\mathbf{y}) - \hat{z}_i \hat{\mathbf{H}}(\mathbf{y})]\mathbf{v}_i = 0$$

for some generalized eigenvector \mathbf{v}_i . Consider now the case of modulated data. Let $\hat{\mathbf{R}}(D(e^{j\phi})\mathbf{y})$ denote the estimate of \mathbf{R} obtained using the modulated data. The ESPRIT estimates $\{\hat{\lambda}_i\}$ of the modulated modes satisfy

$$[\hat{\mathbf{G}}(D(e^{j\phi})\mathbf{y}) - \hat{\lambda}_i \hat{\mathbf{H}}(D(e^{j\phi})\mathbf{y})]\mathbf{w}_i = 0$$

for some generalized eigenvector \mathbf{w}_i . If $\hat{\mathbf{R}}(\mathbf{y})$ is modulation covariant, then

$$\hat{\mathbf{R}}(D(e^{j\phi})\mathbf{y}) = D(e^{j\phi})\hat{\mathbf{R}}(\mathbf{y})D(e^{-j\phi})$$

which implies

$$\begin{aligned} \hat{\mathbf{G}}(D(e^{j\phi})\mathbf{y}) &= e^{j\phi} D(e^{j\phi}) \hat{\mathbf{G}}(\mathbf{y}) D(e^{-j\phi}) \\ \hat{\mathbf{H}}(D(e^{j\phi})\mathbf{y}) &= D(e^{j\phi}) \hat{\mathbf{H}}(\mathbf{y}) D(e^{-j\phi}). \end{aligned}$$

Now, if $\hat{\mathbf{R}}(\mathbf{y})$ is modulation covariant, the estimates $\{\hat{\lambda}_i\}$ satisfy

$$[e^{j\phi} D(e^{j\phi}) \hat{\mathbf{G}}(\mathbf{y}) D(e^{-j\phi}) - \hat{\lambda}_i D(e^{j\phi}) \hat{\mathbf{H}}(\mathbf{y}) D(e^{-j\phi})]\mathbf{w}_i = 0$$

or, equivalently

$$[\hat{\mathbf{G}}(\mathbf{y}) - e^{-j\phi} \hat{\lambda}_i \hat{\mathbf{H}}(\mathbf{y})]D(e^{-j\phi})\mathbf{w}_i = 0.$$

This is just the eigenvalue problem used to obtain the original estimates $\{\hat{z}_i\}$. Clearly, $\hat{\lambda}_i = e^{j\phi} \hat{z}_i$, and this version of the ESPRIT algorithm generates appropriately modulated estimates, provided $\hat{\mathbf{R}}(\mathbf{y})$ is modulation covariant. In [10], it is demonstrated that MUSIC and linear prediction also implicitly rely on modulation covariant estimators of the correlation matrix.

B. Modulation Covariance and Lag Products

In the previous section, we demonstrated that modulation covariance is a desirable property for estimators to possess if the estimated correlation matrix is to be used in a subspace estimation algorithm. In this section, we expand the argument for modulation covariance by showing that modulation covariance is equivalent to requiring that the estimator use only “proper” lagged products of the data.

A sufficient condition for ensuring that an estimator has the quadratic scaling property is to require that the estimator be quadratic in the data. In the remainder of this paper, we will require that all estimators be quadratic in the data. For the single snapshot (or scalar) case, this leads to the following proposition.

Proposition 1: Let $\mathbf{y} = [y_0 \ y_1 \ \dots \ y_{N-1}]^T$ be a vector of complex-valued WSS data. An estimator $\hat{r}_k(\mathbf{y})$ of the correlation coefficient r_k is

- 1) quadratic in the data;
- 2) modulation covariant: $\hat{r}_k(D(e^{j\phi})\mathbf{y}) = e^{jk\phi} \hat{r}_k(\mathbf{y}) \ \forall \mathbf{y} \in \mathbb{C}^N, \phi \in \mathbb{R}$

if and only if $\hat{r}_k(\mathbf{y}) = \mathbf{y} \mathbf{Q}_k \mathbf{y}$, where

$$\begin{aligned} \mathbf{Q}_k &= \begin{bmatrix} x & & & \\ & \ddots & & \\ & & & x \end{bmatrix} (k \geq 0) \\ &= \begin{bmatrix} \text{(nonzero only on} \\ \text{\(k\text{th super-diagonal)} \end{bmatrix} \\ &= \begin{bmatrix} x & & & \\ & \ddots & & \\ & & & x \end{bmatrix} (k < 0) \\ &= \begin{bmatrix} \text{(nonzero only on} \\ \text{\(|k|\text{th sub-diagonal)} \end{bmatrix}. \end{aligned}$$

The proof of this proposition is simple and can be found in [10]. An equivalent form for this condition is

$$\begin{aligned}\hat{r}_k(\mathbf{y}) &= \sum_{i=0}^{N-k-1} q_i y_{i+k} y_i^* \quad (k \geq 0) \\ &= \sum_{i=0}^{N-|k|-1} q_i y_{i+|k|}^* y_i \quad (k < 0).\end{aligned}$$

For example, if we are estimating r_1 , the estimator can only use products with a lag of one: $(y_1 y_0^*, y_2 y_1^*, \dots)$. Hence, the modulation covariance constraint is equivalent to requiring that the estimator use only ‘‘proper’’ lag-products of the data. In the absence of any prior model (structural or parametric) for the correlation sequence, this is a reasonable constraint.

III. PROTOTYPE ESTIMATORS OF CORRELATION

In some instances, it is not necessary to estimate the entire $N \times N$ correlation matrix \mathbf{R} but only the $p \times p$ northwest block. Then, given the data $\mathbf{y} = [y_0 \ y_1 \ \dots \ y_{N-1}]^T$, the objective is to design estimators $\hat{\mathbf{R}}_p(\mathbf{y})$ of the $p \times p$ correlation matrix

$$\mathbf{R}_p = \begin{bmatrix} r_0 & r_{-1} & \dots & r_{1-p} \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{p-1} & \dots & r_1 & r_0 \end{bmatrix}.$$

Again, we will require that the estimator be quadratic in the data and nonnegative definite. The modulation covariance constraint reduces to

$$\begin{aligned}\hat{\mathbf{R}}_p(D(e^{j\phi})\mathbf{y}) &= \mathbf{D}_p(e^{j\phi})\hat{\mathbf{R}}_p(\mathbf{y})\mathbf{D}_p(e^{-j\phi}) \\ \mathbf{D}_p(e^{j\phi}) &= \text{diag}\{1, e^{j\phi}, \dots, e^{j(p-1)\phi}\}.\end{aligned}$$

It will be advantageous to first examine existing estimators of the correlation matrix and determine if they possess the characteristics just enumerated. In addition, we can gain insight into the design of estimators of the correlation matrix by describing the common structure of these estimators.

A. LP Prototypes

Perhaps the most rudimentary, but widely used, estimator of the correlation matrix is the outer product

$$\hat{\mathbf{R}}(\mathbf{y}) = \mathbf{y}\mathbf{y}^*.$$

This estimator serves as our first prototype, and it clearly possesses the three required properties P1–P3. Suppose that the objective is to estimate the $p \times p$ matrix \mathbf{R}_p instead of \mathbf{R} . Every $p \times p$ block on the diagonal of $\hat{\mathbf{R}}(\mathbf{y})$ could serve as an estimator of \mathbf{R}_p . We could then weight each $p \times p$ block and sum the results to construct the estimator

$$\hat{\mathbf{R}}_p(\mathbf{y}) = \sum_{i=0}^{N-p} \alpha_i \mathbf{S}_i \mathbf{y}\mathbf{y}^* \mathbf{S}_i^*.$$

The matrices $\mathbf{S}_i \in \mathbb{R}^{p \times N}$ select each $p \times p$ block on the diagonal of $\hat{\mathbf{R}}(\mathbf{y})$. Some representative examples of \mathbf{S}_i are

given here (blanks denote zeros):

$$\begin{aligned}\mathbf{S}_0 &= \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \ddots \end{bmatrix} \\ \mathbf{S}_1 &= \begin{bmatrix} & 1 & & \\ & & \ddots & \\ & & & 1 & \\ & & & & \ddots \end{bmatrix} \dots \\ \mathbf{S}_{N-p} &= \begin{bmatrix} & & & 1 \\ & & & & \ddots \\ & & & & & 1 \end{bmatrix}.\end{aligned}\quad (2)$$

Because the prototype estimator $\hat{\mathbf{R}}(\mathbf{y}) = \mathbf{y}\mathbf{y}^*$ satisfies our properties, the induced estimator $\hat{\mathbf{R}}_p(\mathbf{y})$ will inherit the quadratic scaling and modulation covariance properties. The nonnegativity constraint will place restrictions on the weight coefficients. Observe that $\hat{\mathbf{R}}_p(\mathbf{y})$ can be written as

$$\begin{aligned}\hat{\mathbf{R}}_p(\mathbf{y}) &= \sum_{i=0}^{N-p} \alpha_i \mathbf{S}_i \mathbf{y}\mathbf{y}^* \mathbf{S}_i^* \\ &= [\mathbf{S}_0 \mathbf{y} \ \dots \ \mathbf{S}_{N-p} \mathbf{y}] \begin{bmatrix} \alpha_0 & & & \\ & \ddots & & \\ & & \alpha_{N-p} & \end{bmatrix} \begin{bmatrix} \mathbf{y}^* \mathbf{S}_0^* \\ \vdots \\ \mathbf{y}^* \mathbf{S}_{N-p}^* \end{bmatrix} \\ &= \mathbf{H}(\mathbf{y}) \mathbf{D} \mathbf{H}^*(\mathbf{y})\end{aligned}\quad (3)$$

where $\mathbf{D} = \text{diag}\{\alpha_0, \alpha_1, \dots, \alpha_{N-p}\}$, and $\mathbf{H}(\mathbf{y})$ is the Hankel data matrix

$$\mathbf{H}(\mathbf{y}) = [\mathbf{S}_0 \mathbf{y} \ \dots \ \mathbf{S}_{N-p} \mathbf{y}] = \begin{bmatrix} y_0 & y_1 & \dots & y_{N-p} \\ y_1 & y_2 & \dots & y_{N-p+1} \\ \vdots & \vdots & & \vdots \\ y_{p-1} & y_p & \dots & y_{N-1} \end{bmatrix}.$$

The formula in (3) clearly shows that $\hat{\mathbf{R}}_p(\mathbf{y}) \geq \mathbf{0} \ \forall \mathbf{y} \in \mathbb{C}^N$ if and only if $\mathbf{D} \geq 0$. Then, each weight coefficient must be nonnegative and can therefore be factored as $\alpha_i = w_i \bar{w}_i$. We now have the two equivalent representations

$$\hat{\mathbf{R}}_p(\mathbf{y}) = \sum_{i=0}^{N-p} \mathbf{W}_i \mathbf{y}\mathbf{y}^* \mathbf{W}_i^* = \mathbf{H}_w(\mathbf{y}) \mathbf{H}_w^*(\mathbf{y}).\quad (4)$$

In these representations, $\mathbf{W}_i = w_i \mathbf{S}_i$ and $\mathbf{H}_w(\mathbf{y}) = [\mathbf{W}_0 \mathbf{y} \ \dots \ \mathbf{W}_{N-p} \mathbf{y}]$. If we set the weight coefficients to $1/N$ in (3), then the resulting estimator

$$\hat{\mathbf{R}}_p(\mathbf{y}) = \frac{1}{N} \mathbf{H}(\mathbf{y}) \mathbf{H}^*(\mathbf{y})$$

is the estimated correlation matrix that arises in the *covariance method of linear prediction*. The other methods of linear prediction also generate estimated correlation matrices that can be represented as in (4). If we augment the set of matrices $\{\mathbf{S}_0, \dots, \mathbf{S}_{N-p}\}$ with the set $\{\mathbf{S}_{-1}, \dots, \mathbf{S}_{1-p}, \mathbf{S}_{N-p+1}, \dots, \mathbf{S}_{N-1}\}$, then we can cast the estimated correlation matrices of the *pre, post, and correlation methods of linear prediction* into the form of (4) [13]. Some

matrix for the data by

$$\mathbf{R} = E\{\mathbf{y}\mathbf{y}^*\} = \begin{bmatrix} r_0 & r_{-1} & \cdots & r_{1-N} \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{N-1} & \cdots & r_1 & r_0 \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{R}_p & * \\ \hline * & * \end{array} \right]$$

where \mathbf{R}_p is the $p \times p$ northwest block of \mathbf{R} . Let $\mathbf{D}_p(e^{j\theta}) = \text{diag}\{1, e^{j\theta}, \dots, e^{j(p-1)\theta}\}$ denote a $p \times p$ modulation matrix. The estimator $\hat{\mathbf{R}}_p(\mathbf{y})$ is

- 1) quadratic in the data;
- 2) nonnegative definite: $\hat{\mathbf{R}}_p(\mathbf{y}) \geq 0 \forall \mathbf{y} \in \mathbb{C}^N$;
- 3) modulation covariant:

$$\begin{aligned} \hat{\mathbf{R}}_p(\mathbf{D}(e^{j\phi})\mathbf{y}) \\ = \mathbf{D}_p(e^{j\phi})\hat{\mathbf{R}}_p(\mathbf{y})\mathbf{D}_p(e^{-j\phi}), \quad \forall \mathbf{y} \in \mathbb{C}^N \quad \text{and} \quad \forall \theta \in \mathbb{R} \end{aligned}$$

if and only if it has a separable representation

$$\hat{\mathbf{R}}_p(\mathbf{y}) = \sum_i \mathbf{W}_i \mathbf{y} \mathbf{y}^* \mathbf{W}_i^*$$

where each $\mathbf{W}_i \in \mathbb{C}^{p \times N}$ is nonzero on only one super- or sub-diagonal.

The theorem we have presented in this section indicates that the most general estimator of the correlation matrix is a multiwindow estimator. Equation (6) indicates that the problem of designing correlation estimators is essentially the dual of spectrum estimator design. In the following, we present an informal argument for using the more general multiwindow estimator structure.

The periodogram, introduced by Schuster [15] in 1897, is a single-window estimator of the power spectrum. In the representation of (5), the periodogram can be written as

$$\hat{S}(e^{j\theta}, \mathbf{y}) = \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{V}^* \mathbf{V} \mathbf{D}(e^{-j\theta}) \mathbf{y}; \quad \mathbf{V} = \frac{1}{\sqrt{N}} [1 \cdots 1].$$

In accordance with the derivation of (6), the periodogram induces the following estimators of the correlation matrix:

$$\hat{\mathbf{R}}(\mathbf{y}) = \mathbf{y}\mathbf{y}^*; \quad \hat{\mathbf{R}}_p(\mathbf{y}) = \frac{1}{N} \mathbf{H}(\mathbf{y}) \mathbf{H}^*(\mathbf{y}).$$

Here, $\mathbf{H}(\mathbf{y})$ is the Hankel data matrix described earlier. These estimators of correlation are widely employed. However, it is known [17] that for many problems, the periodogram is inferior to the more general multiwindow power spectrum estimators. This example suggests, qualitatively, that there is a possibility of improving the performance of algorithms that use only rudimentary estimators. In the sequel, we provide both quantitative and empirical evidence for this assertion.

IV. REPRESENTATION THEOREM FOR TOEPLITZ ESTIMATORS

In this section, we investigate estimators of the correlation matrix that satisfy the constraints of the previous section and that are also required to be Toeplitz. We use the power spectrum estimator

$$\hat{S}(e^{j\theta}, \mathbf{y}) = \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{V}^* \mathbf{V} \mathbf{D}(e^{-j\theta}) \mathbf{y}; \quad \mathbf{V} \in \mathbb{C}^{L \times N} \quad (8)$$

to induce our estimators and then show that under the stated constraints, the estimator must have this form. This idea was mentioned by Thomson in [18].

The power spectrum estimator of (8) can be written as

$$\hat{S}(e^{j\theta}) = \sum_{|k| < N} \hat{r}_k(\mathbf{y}) e^{-jk\theta}$$

where the induced estimators of the correlation coefficients are

$$\hat{r}_k(\mathbf{y}) = \int_{-\pi}^{\pi} \hat{S}(e^{j\theta}) e^{jk\theta} \frac{d\theta}{2\pi}.$$

These estimators have the equivalent forms

$$\begin{aligned} \hat{r}_k(\mathbf{y}) &= \int_{-\pi}^{\pi} \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{V}^* \mathbf{V} \mathbf{D}(e^{-j\theta}) \mathbf{y} e^{jk\theta} \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} \boldsymbol{\psi}^*(e^{-j\theta}) \mathbf{D}_y^* \mathbf{V}^* \mathbf{V} \mathbf{D}_y \boldsymbol{\psi}(e^{-j\theta}) e^{jk\theta} \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} \text{tr}\{\mathbf{D}_y^* \mathbf{V}^* \mathbf{V} \mathbf{D}_y \boldsymbol{\psi}(e^{-j\theta}) \boldsymbol{\psi}^*(e^{-j\theta})\} e^{jk\theta} \frac{d\theta}{2\pi} \quad (9) \end{aligned}$$

where $\boldsymbol{\psi}(e^{-j\theta}) = [1 \ e^{-j\theta} \ \cdots \ e^{-j(N-1)\theta}]^T$, and $\mathbf{D}_y = \text{diag}\{\mathbf{y}\}$. Define the $N \times N$ shift matrix $[\mathbf{Z}_k]_{ij} = \delta(i - j - k)$. Equation (9) reduces to

$$\hat{r}_k(\mathbf{y}) = \text{tr}\{\mathbf{D}_y^* \mathbf{V}^* \mathbf{V} \mathbf{D}_y \mathbf{Z}_k\}. \quad (10)$$

Let the rows of \mathbf{V} be denoted by $\{\mathbf{w}_i^T\}_1^L$, and let $\mathbf{W}_i = \text{diag}\{\mathbf{w}_i\}$. It follows that

$$\begin{aligned} \hat{r}_k(\mathbf{y}) &= \hat{r}_k^T(\mathbf{y}) = \text{tr}\{\mathbf{Z}_k^T \mathbf{D}_y \mathbf{V}^T \bar{\mathbf{V}} \mathbf{D}_y^*\} \\ &= \text{tr}\{\bar{\mathbf{V}} \mathbf{D}_y^* \mathbf{Z}_k^T \mathbf{D}_y \mathbf{V}^T\} \\ &= \sum_{i=1}^L \mathbf{w}_i^* \mathbf{D}_y^* \mathbf{Z}_k^T \mathbf{D}_y \mathbf{w}_i = \sum_{i=1}^L \mathbf{y}^* \mathbf{W}_i^* \mathbf{Z}_k^T \mathbf{W}_i \mathbf{y} \\ &= \mathbf{y}^* \mathbf{Q}_k \mathbf{y} \quad (11) \end{aligned}$$

where $\mathbf{Q}_k = \sum_{i=1}^L \mathbf{W}_i^* \mathbf{Z}_k^T \mathbf{W}_i$. In the following, we address the issue as to whether this estimator is fundamental.

Theorem 2: Let $\mathbf{y} = [y_0 \ y_1 \ \cdots \ y_{N-1}]^T$ be the data sampled from a complex-valued WSS process. Denote the set of correlation coefficients for the process by $\{r_k\}_{1-N}^{N-1}$. The estimators $\{\hat{r}_k(\mathbf{y})\}_{1-p}^{p-1}$ are

- 1) quadratic in the data;
- 2) modulation covariant: $\hat{r}_k(\mathbf{D}(e^{j\phi})\mathbf{y}) = e^{jk\phi} \hat{r}_k(\mathbf{y}) \forall \mathbf{y} \in \mathbb{C}^N, \phi \in \mathbb{R}$;
- 3) they induce nonnegative definite Toeplitz estimators:

$$\hat{\mathbf{R}}_p(\mathbf{y}) = \sum_{|k| < p} \hat{r}_k(\mathbf{y}) \mathbf{Z}_k \geq \mathbf{0}, \quad \forall \mathbf{y} \in \mathbb{C}^N$$

if and only if they can be reduced to the representations

$$\begin{aligned} \hat{r}_k(\mathbf{y}) &= \int_{-\pi}^{\pi} \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{V}^* \mathbf{V} \mathbf{D}(e^{-j\theta}) \mathbf{y} e^{jk\theta} \frac{d\theta}{2\pi} \\ &= \text{tr}\{\mathbf{D}_y^* \mathbf{V}^* \mathbf{V} \mathbf{D}_y \mathbf{Z}_k\} = \mathbf{y}^* \mathbf{Q}_k \mathbf{y} \end{aligned}$$

where

$$\begin{aligned} \mathbf{V} &= \begin{bmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_L^T \end{bmatrix} \in \mathbb{C}^{L \times N}; \quad \mathbf{W}_i = \text{diag}\{\mathbf{w}_i\} \\ \mathbf{D}_y &= \text{diag}\{\mathbf{y}\} \\ \mathbf{Z}_k &= \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} = [\delta(i-j-k)] \\ \mathbf{Q}_k &= \sum_{i=1}^L \mathbf{W}_i^* \mathbf{Z}_k^T \mathbf{W}_i. \end{aligned}$$

Proof: (Sufficiency) Assume $\hat{r}_k(\mathbf{y}) = \text{tr}\{\mathbf{D}_y^* \mathbf{V}^* \mathbf{V} \mathbf{D}_y \mathbf{Z}_k\} = \mathbf{y}^* \mathbf{Q}_k \mathbf{y}$. By construction, \mathbf{Q}_k is nonzero only on its k th diagonal. From Proposition 1, this implies that $\hat{r}_k(\mathbf{y})$ is modulation covariant, and consequently, $\hat{\mathbf{R}}_p(\mathbf{y})$ is modulation covariant. The induced Toeplitz estimator can be written as

$$\begin{aligned} \hat{\mathbf{R}}_p(\mathbf{y}) &= \sum_{|k| < P} \hat{r}_k(\mathbf{y}) \mathbf{Z}_k \\ &= \int_{-\pi}^{\pi} \hat{S}(e^{j\theta}) \boldsymbol{\psi}_p(e^{j\theta}) \boldsymbol{\psi}_p^*(e^{j\theta}) \frac{d\theta}{2\pi}. \end{aligned}$$

Since the power spectrum estimator $\hat{S}(e^{j\theta}, \mathbf{y}) = \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{V}^* \mathbf{V} \mathbf{D}(e^{-j\theta}) \mathbf{y}$ is nonnegative $\forall \mathbf{y} \in \mathbb{C}^N$, the matrix $\hat{\mathbf{R}}_p(\mathbf{y}) \geq \mathbf{0} \forall \mathbf{y} \in \mathbb{C}^N$.

(Necessity) Assume that $\hat{r}_k(\mathbf{y})$ is quadratic in the data, modulation covariant, and induces nonnegative definite Toeplitz matrices. The estimated Toeplitz matrix can be written as

$$\begin{aligned} \hat{\mathbf{R}}_p(\mathbf{y}) &= \int_{-\pi}^{\pi} \left(\sum_{|k| < N} \hat{r}_k(\mathbf{y}) e^{-jk\theta} \right) \boldsymbol{\psi}_p(e^{j\theta}) \boldsymbol{\psi}_p^*(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} \left(\sum_{|k| < N} \mathbf{y}^* \mathbf{Q}_k \mathbf{y} e^{-jk\theta} \right) \boldsymbol{\psi}_p(e^{j\theta}) \boldsymbol{\psi}_p^*(e^{j\theta}) \frac{d\theta}{2\pi}. \end{aligned}$$

Since $\hat{r}_k(\mathbf{y}) = \mathbf{y}^* \mathbf{Q}_k \mathbf{y}$ is modulation covariant

$$e^{-jk\theta} \mathbf{Q}_k = \mathbf{D}(e^{j\theta}) \mathbf{Q}_k \mathbf{D}(e^{-j\theta})$$

which results in

$$\begin{aligned} \hat{\mathbf{R}}_p(\mathbf{y}) &= \int_{-\pi}^{\pi} \left(\sum_{|k| < N} \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{Q}_k \mathbf{D}(e^{-j\theta}) \mathbf{y} \right) \\ &\quad \cdot \boldsymbol{\psi}_p(e^{j\theta}) \boldsymbol{\psi}_p^*(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} (\mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{Q} \mathbf{D}(e^{-j\theta}) \mathbf{y}) \\ &\quad \cdot \boldsymbol{\psi}_p(e^{j\theta}) \boldsymbol{\psi}_p^*(e^{j\theta}) \frac{d\theta}{2\pi} \end{aligned}$$

where $\mathbf{Q} = \sum \mathbf{Q}_k$. It follows that $\hat{\mathbf{R}}(\mathbf{y}) \geq \mathbf{0} \Leftrightarrow \mathbf{Q} \geq \mathbf{0}$. Consequently, \mathbf{Q} can be factored as $\mathbf{Q} = \mathbf{V}^* \mathbf{V}$. The derivation in Section III-B can be used to show that $\hat{r}_k(\mathbf{y})$ has the structure expressed in the statement of the theorem. \square

V. MEAN-SQUARED ERROR BOUND

We have demonstrated that “reasonable” estimators of the correlation sequence must have the structure defined in Theorem 2. This representation allows for the possibility of using multiwindowed copies of the data to construct the estimates of the correlation coefficients. In this section, we establish that the multiwindow estimators of the correlation sequence *may* have smaller mean-squared error than the typical estimator that uses only one window.

The estimated correlation coefficients induce a power spectrum estimator

$$\hat{S} = \sum_{|k| < N} \hat{r}_k(\mathbf{y}) e^{-jk\theta}. \quad (12)$$

This power spectrum estimator is equivalent to the power spectrum estimator described in (8). The error of the power spectrum estimator can be expressed as

$$\begin{aligned} S(e^{j\theta}) - \hat{S}(e^{j\theta}, \mathbf{y}) &= \sum_k r_k e^{-jk\theta} - \sum_{|k| < N} \hat{r}_k(\mathbf{y}) e^{-jk\theta} \\ &= \sum_{|k| < N} [r_k - \hat{r}_k(\mathbf{y})] e^{-jk\theta} + \sum_{|k| \geq N} r_k e^{-jk\theta}. \end{aligned}$$

Assume that the mean-squared error of the power spectrum estimator is integrable. It follows that

$$\begin{aligned} \int_{-\pi}^{\pi} \text{mse}\{\hat{S}(e^{j\theta}, \mathbf{y})\} \frac{d\theta}{2\pi} &= E \left\{ \int_{-\pi}^{\pi} |S(e^{j\theta}) - \hat{S}(e^{j\theta}, \mathbf{y})|^2 \frac{d\theta}{2\pi} \right\} \\ &= E \left\{ \sum_k |r_k - \hat{r}_k(\mathbf{y})|^2 \right\} \\ &= E \left\{ \sum_{|k| < N} |r_k - \hat{r}_k(\mathbf{y})|^2 \right\} + \sum_{|k| \geq N} |r_k|^2 \quad (13) \\ &= \sum_{|k| < N} \text{mse}\{\hat{r}_k(\mathbf{y})\} + \sum_{|k| \geq N} |r_k|^2. \quad (14) \end{aligned}$$

Mullis and Scharf [11] have established a lower bound on the mean-squared error of a power spectrum estimator with L windows

$$\text{mse}\{\hat{S}(e^{j\theta}, \mathbf{y})\} \geq \frac{S^2(e^{j\theta})}{L+1}. \quad (15)$$

Then, if the power spectrum is square-integrable, we have

$$\begin{aligned} \sum_{|k| < N} \text{mse}\{\hat{r}_k(\mathbf{y})\} &= \int_{-\pi}^{\pi} \text{mse}\{\hat{S}(e^{j\theta}, \mathbf{y})\} \frac{d\theta}{2\pi} - \sum_{|k| \geq N} |r_k|^2 \\ &\geq \frac{1}{L+1} \int_{-\pi}^{\pi} S^2(e^{j\theta}) \frac{d\theta}{2\pi} - \sum_{|k| \geq N} |r_k|^2 \\ &= \frac{1}{L+1} \sum_{|k| < N} |r_k|^2 - \frac{L}{L+1} \sum_{|k| \geq N} |r_k|^2. \quad (16) \end{aligned}$$

As was pointed out by a reviewer, the bound of (16) is not guaranteed to be positive. In the following, we argue that despite this fact, the bound still provides a valid argument for using multiwindow estimators of correlation.

In the derivation of (16), it was assumed that the power spectrum was square integrable. Consequently, the infinite series $\sum_k |r_k|^2$ converges. This implies that there exists an $N(L)$ such that for $N \geq N(L)$, the bound in (16) is positive. If, for a fixed N , the bound in (16) is negative, then this is an indication that too many windows are being used. Similarly, for fixed L , $N(L)$ can be interpreted as a lower bound on the size of the data vector for which an L window estimator is appropriate. In the following, we argue that the bound of (16) will be positive in those situations for which a multiwindow estimator is applicable.

Mullis and Scharf [11] show that the bound on the mean-squared error of the power spectrum estimator (15) is tight only if the power spectrum is “essentially constant” on intervals of length $2\pi L/N = 2\pi\beta$. To quantify this statement, we make the approximation

$$\begin{aligned} \beta S(e^{j\theta}) &\approx \int_{\theta-\beta\pi}^{\theta+\beta\pi} S(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= \sum_k r_k \int_{\theta-\beta\pi}^{\theta+\beta\pi} e^{-jk\theta} \frac{d\theta}{2\pi} \\ &= \beta \sum_k r_k \text{sinc}(k\beta\pi) e^{-jk\theta}. \end{aligned} \quad (17)$$

The total squared error in this approximation is

$$\begin{aligned} \epsilon &= \int_{-\pi}^{\pi} |S(e^{j\theta}) - \sum_k r_k \text{sinc}(k\beta\pi) e^{-jk\theta}|^2 \frac{d\theta}{2\pi} \\ &= \sum_k |r_k|^2 (1 - \text{sinc}(k\beta\pi))^2. \end{aligned} \quad (18)$$

We say that a power spectrum is “essentially constant” on intervals of length $2\pi\beta$ if ϵ of (18) is small. Let $\beta = L/N$, and note that

$$\begin{aligned} \epsilon &= \sum_k |r_k|^2 (1 - \text{sinc}(kL\pi/N))^2 \\ &\geq \sum_{|k| \geq N} |r_k|^2 (1 - \text{sinc}(kL\pi/N))^2 \\ &\geq \frac{L}{L+1} \sum_{|k| \geq N} |r_k|^2. \end{aligned} \quad (19)$$

The result of (19) implies that if the power spectrum is essentially constant on intervals of length $2\pi L/N$, then the negative term in the MSE bound of (16) is bounded by a small number.

This observation places a restriction on the number of windows that should be used in the construction of multiwindow correlation estimators. If the power spectrum is essentially constant on intervals of length $2\beta\pi$, then the number of windows should be approximately $L = N\beta$. Time series with a white noise plus lines spectra are not amenable to multiwindow estimators of correlation. Conversely, the multiwindow approach is appropriate if the process has a

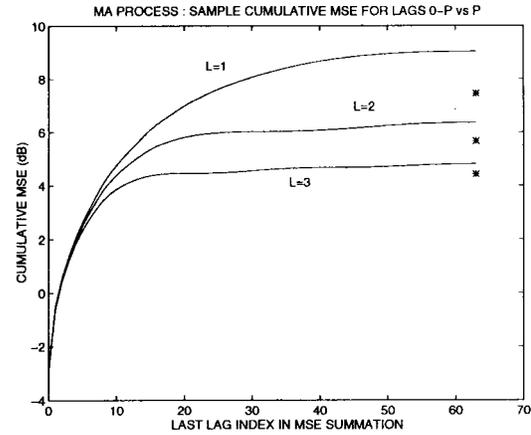


Fig. 1. Sample cumulative MSE versus correlation lag for a MA process.

relatively smooth power spectrum in comparison with the resolution limit implied by the number of data samples.

Equation (15) indicates that a window that is designed to produce a good power spectrum estimator will be a “good” window for estimators of the correlation sequence. In [11], it is argued that a “good” set of windows $\{\mathbf{w}_i^T\}_1^L$ is composed of orthogonal vectors, i.e., $\mathbf{w}_i^T \mathbf{w}_j = 0$ for $i \neq j$. In addition, those windows that concentrate their energy $|W(e^{j\theta})|^2 = |\mathbf{w}^T \psi_N(e^{-j\theta})|^2$ around baseband are superior to those that do not have this lowpass property. In the examples that follow, Slepian (prolate spheroidal) windows are utilized. A description of how to construct these windows can be found in [11].

VI. EXAMPLES

A. MA Process

The data of Fig. 1 were generated as follows. A time series of length $N = 64$ was generated by passing unit variance white noise through the MA filter $B(z) = (1 - .8e^{-j\pi/4}z^{-1})(1 - .8e^{j\pi/4}z^{-1})$. For each time series realization, the estimates of $\{r_k\}_0^{N-1}$ were generated using the estimators described in Theorem 2, with $L = 1, 2$, and 3 Slepian windows. The sample cumulative MSE's $\sum_0^p \text{mse}(\hat{r}_k(\mathbf{y}))$ for $p = 0, \dots, N-1$ were computed from 1000 independent trials. These curves indicate that the multiwindow estimators have better mean-squared error performance than the single window estimator in this scenario. The bulk of the performance gain is attributable to the multiwindow estimator's lower MSE in estimating the higher order correlation coefficients. The asterisks in Fig. 1 are the bounds of (16) for this MA process.

B. AR Process

The data of Fig. 2 were generated as were those in Fig. 1, except that the time series was an AR process generated by passing white noise through the AR filter $1/B(z)$. The polynomial $B(z)$ was the same as above. Again, for this scenario, the multiwindow estimators exhibited better mean-squared error performance than the single window estimator.

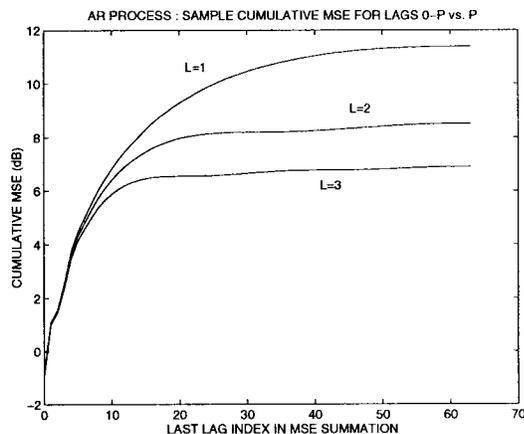


Fig. 2. Sample cumulative MSE versus correlation lag for an AR process.

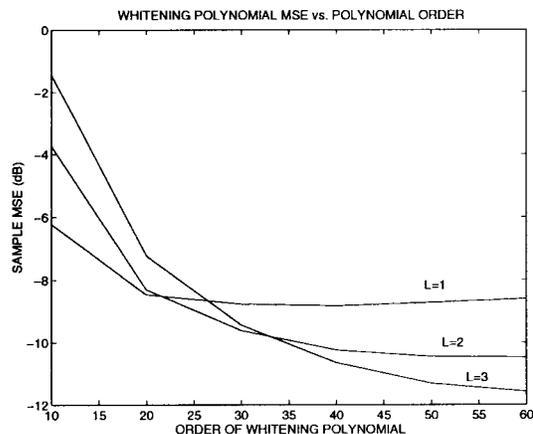


Fig. 3. Sample whitening polynomial MSE versus data length.

C. Linear Prediction

The data of Fig. 3 were generated as follows. An MA time series of variable length N was constructed as in Example 6.1. An estimate of the $N \times N$ correlation matrix $\hat{\mathbf{R}}$ was obtained using the Toeplitz estimators of Theorem 2. Again, $L = 1, 2$, and 3 Slepian windows were used. For each Monte-Carlo trial, and for each N , an estimate of the unit-norm whitening polynomial coefficients $\hat{\mathbf{a}}$ was generated using

$$\hat{\mathbf{a}} = \frac{\hat{\mathbf{R}}^{-1} \boldsymbol{\delta}}{(\boldsymbol{\delta}^T \hat{\mathbf{R}}^{-2} \boldsymbol{\delta})^{1/2}}.$$

The sample MSE of $\hat{\mathbf{a}}$ was computed from 1000 independent trials for $N = 10, \dots, 60$. The results are plotted in Fig. 3. This figure illustrates the bias (resolution) versus variance tradeoff inherent in multiwindow estimators. For small data records, the multiwindow estimators perform poorly in comparison to the single window estimator. In this case, the bias introduced by the multiwindow estimators dominates any reduction in variance. Stated another way, the spectrum of the process cannot be considered “essentially constant” on intervals of length $2\pi L/N$ for small values of N . For larger values of N , the multiwindow estimators’ performance improves and is attributable to the variance reduction inherent in a multiwindow estimator dominating any increase in bias.

VII. CONCLUSIONS

In this paper, we have presented a representation for estimators of the correlation matrix that are required to be quadratic in the data, nonnegative definite, and modulation covariant. We have also derived a representation for estimators that have the additional constraint that the estimated matrix be Toeplitz. These estimators admit structures that use multiply windowed copies of the data. We have demonstrated that many commonly used estimators are the single window duals of the periodogram power spectrum estimator. This observation, in conjunction with the mean-squared error bound and the example we have presented, indicates that many algorithms that use estimators of correlation have the *potential* to be improved by using the multiwindow estimators expressed in

the theorems. The examples and discussions we have provided indicate those situations in which a multiwindow estimator should be used.

REFERENCES

- [1] G. Bienvenu and L. Kopp, “Adaptivity to background noise spatial coherence for high resolution passive methods,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Denver, CO, Apr. 1980, pp. 307–310.
- [2] R. B. Blackman and J. W. Tukey, “The measurement of power spectra from the point of view of communications engineering,” *Bell Syst. Tech. J.*, vol. 33, pp. 185–282; pp. 485–569, 1958.
- [3] M. P. Clark and L. L. Scharf, “Frequency-wavenumber spectrum analysis using quadratic estimators,” *Proc. IEEE Workshop Stat. Signal Array Process.*, Victoria, B.C., Canada, 1992.
- [4] H. Clergeot, “Choix entre differentes methodes quadratiques d’estimation spectrale,” *Ann. Telecommun.*, vol. 39, no. 3, pp. 113–128, 1984.
- [5] P. J. Daniell, “Discussion on the symposium of autocorrelation in time series,” *J. R. Stat. Soc. (Suppl.)*, vol. 8, pp. 88–90, 1946.
- [6] U. Grenander and M. Rosenblatt, *Statistical Analysis of Time Series*. New York: Wiley, 1957.
- [7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
- [8] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [9] L. T. McWhorter and L. L. Scharf, “Multiwindow estimators of correlation,” in *Proc. 23rd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1994.
- [10] L. T. McWhorter, “Representations for covariance bounds and time series estimators,” Ph.D. dissertation, Univ. Colorado, Boulder, 1994.
- [11] C. T. Mullis and L. L. Scharf, “Quadratic estimators of the power spectrum,” in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1991, ch. 1.
- [12] R. Roy and T. Kailath, “ESPRIT-Estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July, 1989.
- [13] L. L. Scharf and L. T. McWhorter, “Quadratic estimators of the correlation matrix,” in *Proc. Acoust., Speech, Signal Processing Workshop Stat. Signal Array Process.*, Quebec City, P.Q., Canada, June 27–29, 1994.
- [14] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” in *Proc. RADAR Spectral Estimation Workshop*, Rome, NY, 1979, pp. 243–258.
- [15] A. Schuster, “On lunar and solar periodicities of earthquakes,” *Proc. R. Soc.*, vol. 61, pp. 455–465, 1897.
- [16] P. Stoica and A. Nehorai, “MUSIC, maximum likelihood, and the Cramer-Rao bound,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
- [17] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proc. IEEE*, vol. 70, pp. 1055–1096, Sept. 1982.
- [18] ———, “Quadratic-inverse spectrum estimates: applications to paleoclimatology,” *Philos. Trans. R. Soc. Lond.*, vol. 332, no. 1627, p. 539, Sept. 1990.



L. Todd McWhorter received the B.S. degree in electrical engineering from Colorado State University, Fort Collins, in 1984. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Colorado, Boulder, in 1991 and 1994, respectively.

From 1985 to 1989, he was employed by Woodward Governor Company, Fort Collins, as an applications engineer and circuit designer. He served as a research associate and instructor at the University of Colorado from 1989 to 1996. He is currently with Orincon Corporation, San Diego, CA. His current research interests include time series and array processing algorithms, multiwindow estimators of second- and higher order statistics, covariance bound theory, and spectral estimation.



Louis L. Scharf (F'86) received the Ph.D. degree in electrical engineering in 1969 from the University of Washington, Seattle.

From 1969 to 1971, he was a member of the Technical Staff at Honeywell's Marine Systems Center, Seattle. He served as Professor of Electrical Engineering and Statistics at Colorado State University, Fort Collins, from 1971 to 1981. From 1982 to 1985, he was Professor and Chair of Electrical and Computer Engineering at the University of Rhode Island, Kingston. He is currently Professor of Electrical and Computer Engineering at the University of Colorado, Boulder, where he teaches and conducts research in signal processing. In 1974, he was Visiting Associate Professor at Duke University, Durham, NC. In 1977, he was a member of the technical staff in the CNRS Laboratoire des Signaux et Systemes, Gif-sur-Yvette, France, and Professeur Associe at the University of South Paris, Orsay, France. In 1981, he was a Visiting Professor at Ecole Nationale Supérieure des Telecommunications, Paris, France, and at the University of La Plata, Buenos Aires, Argentina. He was a Visiting Professor at Institut Eurecom, Sophia-Antipolis, France, in 1992.

Professor Scharf is a past member of the ASSP ADCOM. He has served on the Editorial Board of Signal Processing and is a past Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was Technical Program Chairman of the IEEE International Conference on Acoustics, Speech, and Signal Processing in 1980. He currently serves on the Asilomar Conference Board. In 1994, he served as a Distinguished Lecturer for the IEEE Signal Processing Society, and in 1995, he received the Society's Technical Achievement Award.