$$(S_n)_{ij} \equiv \sum_{t=1}^{N-i-j} X_{t+i}X_{t+j}^* \qquad 0 \le i, j \le n. \tag{6}$$

Equation (6) is significant since we now have Ω_n as an explicit function of the components of $\boldsymbol{a}$ (contained in the vector $\boldsymbol{\alpha}$). Though not Toeplitz, S_n possesses the important *double* Hermitian property since $(S_n)_{ij} = (S_n)_{ji}^* = (S_n)_{n-j,n-i}^*$.

We continue with the maximization by combining (4) and (5)

$$\mathcal{L}_n = -N \log\left(\frac{1}{N}\Omega_n\right) + n \log\left(1 - |k_n|^2\right) + \log|\boldsymbol{R}_{f,n-1}^{-1}|. \tag{7}$$

We define the real and imaginary parts of the reflection coefficient k_n as ξ_n and η_n and set partial derivatives to zero

$$\frac{\partial \mathcal{L}_n}{\partial \xi_n} = -\frac{N}{\Omega_n}\frac{\partial \Omega_n}{\partial \xi_n} - \frac{2n\xi_n}{1 - (\xi_n^2 + \eta_n^2)} = 0$$

$$\frac{\partial \mathcal{L}_n}{\partial \eta_n} = -\frac{N}{\Omega_n}\frac{\partial \Omega_n}{\partial \eta_n} - \frac{2n\eta_n}{1 - (\xi_n^2 + \eta_n^2)} = 0. \tag{8}$$

Computing the derivatives of Ω_n explicitly, we find

$$\frac{\partial \Omega_n}{\partial \xi_n} = 2\,\mathrm{Re}\begin{pmatrix} 0 \\ \boldsymbol{J}\boldsymbol{\alpha}_{n-1}^* \end{pmatrix}^{\dagger} S_n \alpha_n$$

$$\frac{\partial \Omega_n}{\partial \eta_n} = 2\,\mathrm{Im}\begin{pmatrix} 0 \\ \boldsymbol{J}\boldsymbol{\alpha}_{n-1}^* \end{pmatrix}^{\dagger} S_n \alpha_n. \tag{9}$$

The notations "Re" and "Im" stand for real and imaginary parts, respectively. The matrix $\boldsymbol{J}$ is the rank n *reverse operator* with ones on the secondary diagonal and zeroes elsewhere [3]. For example, the order 3 matrix $\boldsymbol{J}$ is

$$\boldsymbol{J} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Equations (9) employed the identity

$$\boldsymbol{\alpha}_n = \begin{pmatrix} \boldsymbol{\alpha}_{n-1} \\ 0 \end{pmatrix} + (\xi_n + i\eta_n)\begin{pmatrix} 0 \\ \boldsymbol{J}\boldsymbol{\alpha}_{n-1}^* \end{pmatrix}.$$

We now solve (8) for the real and imaginary parts of the reflection coefficient at step n (ξ_n and η_n). Multiplying the first of equations (8) by η_n and the second by ξ_n and subtracting the two shows that the quantity $k_n^* \Gamma$ has zero imaginary part where we define Γ (and other quantities) as

$$\Gamma \equiv \begin{pmatrix} 0 \\ \boldsymbol{J}\boldsymbol{\alpha}_{n-1}^* \end{pmatrix}^{\dagger} S_n \begin{pmatrix} \boldsymbol{\alpha}_{n-1} \\ 0 \end{pmatrix}$$

$$\Delta \equiv \begin{pmatrix} 0 \\ \boldsymbol{J}\boldsymbol{\alpha}_{n-1}^* \end{pmatrix}^{\dagger} S_n \begin{pmatrix} 0 \\ \boldsymbol{J}\boldsymbol{\alpha}_{n-1}^* \end{pmatrix}$$

$$\Psi \equiv \begin{pmatrix} \boldsymbol{\alpha}_{n-1} \\ 0 \end{pmatrix}^{\dagger} S_n \begin{pmatrix} \boldsymbol{\alpha}_{n-1} \\ 0 \end{pmatrix}. \tag{10}$$

Note that Γ is complex while Δ and Ψ are real. Since we know $k_n^* \Gamma$ to be real, we may define the real quantity ω and, with no loss of generality, state that

$$k_n = \omega \Gamma / |\Gamma|. \tag{11}$$

The double Hermitian property of S_n and the definition of $\boldsymbol{J}$ imply that $\Psi = \Delta$.

We now return to (8) with this new expression for k_n. Multiplying the first of equations (8) by ξ_n and the second by η_n and adding the two yields the cubic equation $Q(\omega) = 0$. More specifically

$$\begin{aligned} Q(\omega) &= -\Delta(N-n)\omega^3 + |\Delta|(2n-N)\omega^2 \\ &\quad + \Delta(N+n)\omega + N|\Gamma| \\ &= n\omega(\Delta + 2|\Gamma|\omega + \Delta\omega^2) + N(1-\omega^2)(|\Gamma| + \Delta\omega) \\ &= 0. \end{aligned} \tag{12}$$

Thus, we have reduced the determination of ω, and hence the complex reflection coefficient k_n, to solution of a real cubic equation. In this sense the complex RMLE algorithm is quite similar to Kay's real version [1].

Since the magnitude of the reflection coefficient k_n equals the magnitude of ω, stability of this complex recursive maximum likelihood estimation procedure requires $|\omega| < 1$. We would like to be able to show, therefore, that there always exists a root of the cubic polynomial $Q(\omega)$ with magnitude less than unity. We have failed to do this directly. But we can argue that the solution ω, and hence k_n, must have this property by inspection of (7). Clearly the quantity to be maximized ($\mathcal{L}_n$) decreases to negative infinity as $|k_n|$ approaches unity due to the $\log(1 - |k_n|^2)$ term. Furthermore, even though $\Omega_n (\equiv \boldsymbol{X}^{\dagger}\boldsymbol{R}_{fn}^{-1}\boldsymbol{X})$ is also a function of k_n, we know that Ω_n is positive since $\boldsymbol{R}_{fn}$, and hence $\boldsymbol{R}_{fn}^{-1}$, is positive definite. The logarithm of $N^{-1}\Omega_n$ in (7) thus remains finite. Given that $\mathcal{L}_n$ is real and finite within the unit circle in the complex plane ($|k_n| < 1$) and approaches negative infinity on the unit circle, $\mathcal{L}_n$ must attain a maximum value for $|k_n| < 1$.

III. Summary

We have extended the recursive maximum likelihood estimation algorithm conceived by Kay [1] to complex data sets. The complex version requires the same level of computation as that for real data. We have argued, without direct proof, that the algorithm is stable in the sense that the magnitude of the reflection coefficient at each step is less than unity.

References

[1] S. M. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, p. 56, 1983; also S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[2] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford University, Stanford, CA, 1975.

[3] J. Makhoul, "On the eigenvectors of symmetric Toeplitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, p. 868, 1981.

Estimation of Structured Covariance Matrices and Multiple Window Spectrum Analysis

BARRY D. VAN VEEN AND LOUIS L. SCHARF

***Abstract*—An intimate relationship between low rank modeling and multiple window spectrum estimation is demonstrated by using maxi-**

Manuscript received January 30, 1988; revised September 30, 1989. This work was supported in part by the Wisconsin Alumni Research Foundation, the National Science Foundation under MIP-8958559, and by the Office of Naval Research, Statistics and Probability Branch, under Contract N00014-85-K-0256.

B. D. Van Veen is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706-1691.

L. L. Scharf is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425.

IEEE Log Number 9036497.

mum likelihood estimates of structured covariance matrices. The power in a narrow spectral band is estimated by estimating the variances in a low rank signal plus noise covariance model. This model is swept through the entire frequency band to obtain an estimate of power as a function of frequency. The resulting spectrum estimates are given by weighted combinations of eigenspectra. Each eigenspectrum results from projecting the data onto an orthogonal component of the signal subspace and squaring. The multiple window spectrum estimates of Thomson correspond to a particular choice for the low rank signal model. The low rank modeling and structured covariance matrix framework is also used to derive the maximum likelihood estimate for the center frequency of a signal in noise. This estimate is also obtained from a weighted combination of eigenspectra.

I. Introduction

This correspondence illustrates the relationship between multiple window spectrum analysis [1] and maximum likelihood (ML) estimation of structured covariance matrices [2]–[7]. The approach is based on low rank models for the covariance matrix corresponding to narrow-band signals. As a special case, our results reproduce those of Thomson [1]. The unifying theme is that data may be projected onto subspaces and manipulated in each subspace to obtain the spectrum estimate. This view shows that spectrum estimation fits into the framework proposed in [8] for low rank modeling of stationary random vectors.

Structured covariance matrices have been studied in the statistical literature [3]–[7] in the context of factor analysis and analysis of variance. Estimates of structured covariance matrices are utilized in the context of spectral estimation in [2] and [9]. In [2], a Toeplitz constrained covariance matrix is estimated from the data and used to estimate a maximum entropy spectrum. Toeplitz and rank constraints are enforced in [9] and the estimated covariance matrix is used in the MUSIC algorithm. Both of these represent examples where constrained covariance estimates are employed within spectrum estimation algorithms.

Estimates of a structured covariance matrix are used to directly estimate spectra in [10] and [11]. This is accomplished by estimating the variances associated with the components of distinct, densely sampled Fourier models, each representing a part of the Nyquist band. Here we replace the line spectrum model of [10] and [11] with a continuous narrow-band model in order to estimate the spectrum from the variance in an orthogonal, low rank, signal-plus-noise model. The low rank model, representing a narrow frequency band of the spectrum is swept throughout the entire frequency band to obtain variance estimates as a function of frequency. Reduced rank modeling is a general principle for exchanging model bias for model variance [8].

The work presented here clearly illustrates the connection between rank reduction and the multiple window spectrum analysis techniques proposed by Thomson [1].[1] The multiple window approach estimates the spectrum using weighted combinations of eigenspectra, defined in [1] as the Fourier spectra obtained by windowing the data with prolate spheroidal wave functions of increasing order. Eigenspectra are defined here to be the norm of the data after it has been projected onto a component of the low rank signal space. We derive the ML estimate of the center frequency of a bandpass signal in noise and show that it is based on a weighted combination of eigenspectra. Although not further discussed in this paper, we note that there is a clear connection between the problems of estimating power in a narrow-band and of detecting a narrow-band signal. Using the results of [12] (see [12, eq. (9)]) it is easy to show that the likelihood ratio for detection of a low rank signal in white noise is also given by a weighted combination of eigenspectra.

The paper is outlined as follows. Section II reviews maximum likelihood estimation of structured covariance matrices, low rank modeling of narrow-band signals, and multiple window spectrum analysis. The structured covariance approach to spectrum analysis is developed in Section III and its relationship to multiple window spectrum analysis is discussed. In Section IV, low rank models and structured covariance matrix estimates are used to estimate the center frequency of a bandpass signal in noise. A summary is given in Section V.

[1]The relationship between rank reduction and multiple window spectrum analysis has recently been discussed in the context of quadratic estimators of power spectra in [17].

II. Background

The joint probability density function of a set of M independent vector samples $\boldsymbol{x}_m$, $1 \le m \le M$, drawn from a complex N-dimensional zero mean Gaussian process ($\boldsymbol{x}_m \sim N(\boldsymbol{0}, \boldsymbol{R})$) is

$$p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M) = \pi^{-MN}|\boldsymbol{R}|^{-M} \exp\left[-M \operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{S})\right] \quad (1)$$

where $\boldsymbol{R}$ is the true covariance matrix and $\boldsymbol{S}$ is the sample covariance matrix

$$\boldsymbol{S} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{x}_m \boldsymbol{x}_m^H.$$

Let $|\boldsymbol{A}|$ denote the determinant of $\boldsymbol{A}$ and tr $\boldsymbol{A}$ denote the trace of $\boldsymbol{A}$. Boldface is used to denote matrix and vector quantities, superscripts T and H denote matrix transpose and conjugate transpose, respectively.

A. Estimation of Structured Covariance Matrices

Let $\boldsymbol{R}(\boldsymbol{\theta})$ denote the covariance matrix of specified structure as a function of the real parameter vector, $\boldsymbol{\theta}$. The maximum likelihood estimation problem is to determine $\boldsymbol{\theta}$ so that (1) is maximized when $\boldsymbol{R}$ is replaced by $\boldsymbol{R}(\boldsymbol{\theta})$. Taking the natural logarithm we write the maximum likelihood problem as

$$\max_{\boldsymbol{\theta}} -\ln|\boldsymbol{R}(\boldsymbol{\theta})| - \operatorname{tr}\left[\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{S}\right]. \quad (2)$$

In general, this maximization requires solution of a set of nonlinear equations, but for some models an explicit solution can be obtained. Uniqueness and existence of solutions is addressed in [2] and [13]. In this paper we represent $\boldsymbol{R}(\boldsymbol{\theta})$ by the linear model used in [5]–[7], [10], and [11]:

$$\boldsymbol{R}(\theta) = \sum_{i=1}^{p} \theta_i \boldsymbol{Q}_i. \quad (3)$$

Explicit solutions to (2) are obtained because of the form of the $\boldsymbol{Q}_i$.

B. Low Rank Modeling of Bandpass Signals

Let $\boldsymbol{x}$ represent data from a process which has all of its energy concentrated in the spectral band $-W \le f \le W$. A low rank model for $\boldsymbol{x}$ is $\hat{\boldsymbol{x}} = \boldsymbol{U}\boldsymbol{\beta}$ where $\boldsymbol{U}$ is an N by p matrix having orthonormal columns ($\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}$) and $\boldsymbol{\beta}$ is a p by 1 vector of coefficients. The mean square modeling error is

$$e^2 = E\left\{|\boldsymbol{x} - \hat{\boldsymbol{x}}|^2\right\} = \operatorname{tr}\boldsymbol{R} - \operatorname{tr}\boldsymbol{U}^H\boldsymbol{R}\boldsymbol{U} \quad (4)$$

where $\boldsymbol{R} = E\{\boldsymbol{x}\boldsymbol{x}^H\}$. It is well known that the error is minimized by choosing the columns of $\boldsymbol{U}$ as the eigenvectors of $\boldsymbol{R}$ corresponding to the p largest eigenvalues. Ordering the eigenvalues of $\boldsymbol{R}$, λ_i, from largest to smallest, the minimum mean-square error is given by

$$e^2_{\min} = \sum_{i=p+1}^{N} \lambda_i.$$

Thus, the low rank model provides an accurate representation of $\boldsymbol{x}$ if the $N - p$ smallest eigenvalues of $\boldsymbol{R}$ are approximately zero.

Denote the covariance matrix corresponding to a unit power white process on $-W \le f \le W$ as $\boldsymbol{R}_w$. Slepian [14] has shown that for reasonably large N the first $2NW$ eigenvalues of $\boldsymbol{R}_w$ are close to unity and the remainder are approximately zero. This suggests that a bandpass white process is well modeled with $p = 2NW$ since $e^2_{\min} \approx 0$. The eigenvectors of $\boldsymbol{R}_w$ are termed discrete prolate spheroidal sequences (DPSS).

Suppose the bandpass process has power spectrum $S(f)$. Letting $\boldsymbol{d}(f) = [1\ e^{j2\pi f} e^{j4\pi f} \cdots e^{j(N-1)2\pi f}]^H$, $\boldsymbol{R}$ is expressed in terms of $S(f)$ as

$$\boldsymbol{R} = \int_{-W}^{W} S(f)\boldsymbol{d}(f)\boldsymbol{d}^H(f)\, df. \tag{5}$$

Expressing (4) in terms of (5) we obtain

$$e^2 = \int_{-W}^{W} S(f)[\text{tr}\, \boldsymbol{d}(f)\boldsymbol{d}^H(f) - \text{tr}\, \boldsymbol{U}^H\boldsymbol{d}(f)\boldsymbol{d}^H(f)\boldsymbol{U}]\, df. \tag{6}$$

The DPSS minimize e^2 for $S(f) = 1$. If $p = 2NW$, then the bracketed term in the integrand is approximately zero for all $-W \le f \le W$. Thus, the DPSS provide a good low rank model for bandpass signals even if $S(f)$ is not white.

C. Multiple Window Spectrum Analysis

Multiple window spectrum estimates are derived in [1] using a weighted eigenfunction expansion to obtain an approximate solution to the fundamental equation of spectrum estimation

$$y(f) = \int_{-1/2}^{1/2} \frac{\sin N\pi(f-v)}{\sin \pi(f-v)}\, dZ(v). \tag{7}$$

$y(f)$ is the discrete Fourier transform of the N point data sequence $x(n)$ and $dZ(f)$ is a zero mean orthogonal increment process related to the spectrum $S(f)$ by $S(f)\, df = E\{|dZ(f)|^2\}$. This equation is "solved" in a local region $(f_0 - W, f_0 + W)$ about some frequency of interest, f_0.

A solution for $S(f)$ of the form

$$\hat{S}(f) = \sum_{k=1}^{K} w_k(f)\left|y_k(f)\right|^2 \tag{8}$$

is proposed in [1]. The kth eigenspectrum $|y_k(f)|^2$ is given by

$$\left|y_k(f)\right|^2 = \left|\sum_{n=0}^{N-1} x(n)v_k(n)e^{-j2\pi fn}\right|^2. \tag{9}$$

Thus each eigenspectrum is the magnitude squared of a windowed $(v_k(n))$ discrete Fourier transform of the data. Thomson chooses the DPSS as the window sequences and suggests several methods for determining the weights $w_k(f)$. Note that the multiple window spectrum analysis technique consists of combining the spectra obtained by applying different windows to the data.

III. Structured Covariance Matrix Analysis with Frequency Sweeping

We view spectrum analysis as the problem of determining the power present in narrow frequency bands given a finite number of observations from a process. This is consistent with historical approaches to the subject and with Thomson's perspective [1]. In this section we employ maximum likelihood estimates of structured covariance matrices to estimate power in narrow spectral bands. Power is estimated from the variances associated with signal and noise components in a low rank signal plus noise model. Several low rank signal and noise models corresponding to different covariance structure assumptions are studied. The correspondence to multiple window spectrum analysis is illustrated and the section is concluded with a discussion.

A. Fixed Covariance Matrix Models

This subsection presents the model and derives the maximum likelihood estimates of the model parameters for covariance matrices corresponding to fixed frequency bands. In the following subsection these models are swept through the entire frequency band to obtain spectrum estimates.

Assume the observed data vectors, $\boldsymbol{x}_m 1 \le m \le M$, are modeled by

$$\hat{\boldsymbol{x}}_m = \boldsymbol{U}\boldsymbol{\beta}_m + \boldsymbol{n}_m \tag{10}$$

where $\boldsymbol{U}$ is an N by $p(p < N)$ matrix satisfying $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}$. $\boldsymbol{U}\boldsymbol{\beta}_m$ represents a low rank model for the signal and $\boldsymbol{n}_m$ represents noise where noise is defined as any data component not represented by the low rank signal model. If $\boldsymbol{U}$ represents all the energy within a frequency band, then noise corresponds to energy exterior to the band. Define an N by $N - p$ matrix $\boldsymbol{V}$ such that $\boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}$ and $\boldsymbol{V}^H\boldsymbol{U} = \boldsymbol{0}$. $\boldsymbol{V}$ provides a low rank model for the noise: $\boldsymbol{n}_m = \boldsymbol{V}\gamma_m$. $\boldsymbol{\beta}_m$ and γ_m are assumed to be samples from p and $N - p$ dimensional uncorrelated ($E\{\boldsymbol{\beta}_m\gamma_m^H\} = \boldsymbol{0}$), zero mean Gaussian random processes. Let $E\{\boldsymbol{\beta}_m\boldsymbol{\beta}_m^H\} = \boldsymbol{B}$ and $E\{\gamma_m\gamma_m^H\} = \boldsymbol{G}$. The data covariance, $\boldsymbol{R} = E\{\boldsymbol{x}_m\boldsymbol{x}_m^H\}$, is thus modeled by

$$\hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H + \boldsymbol{V}\boldsymbol{G}\boldsymbol{V}^H \tag{11}$$

where $\boldsymbol{B}$ and $\boldsymbol{G}$ are unknown and are estimated from the data. The power associated with the signal is tr $\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H$ = tr $\boldsymbol{B}$ while the noise power is tr $\boldsymbol{G}$.

Let $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{G}}$ denote the ML estimates of $\boldsymbol{B}$ and $\boldsymbol{G}$. They are obtained from (2) as follows. The orthogonality of the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ implies that

$$\det \hat{\boldsymbol{R}} = \det \begin{bmatrix} \boldsymbol{B} & 0 \\ 0 & \boldsymbol{G} \end{bmatrix} = \det \boldsymbol{G} \det \boldsymbol{B} \tag{12}$$

and

$$\text{tr}\,[\hat{\boldsymbol{R}}^{-1}\boldsymbol{S}] = \text{tr}\,[\boldsymbol{U}\boldsymbol{B}^{-1}\boldsymbol{U}^H\boldsymbol{S}] + \text{tr}\,[\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{V}^H\boldsymbol{S}]. \tag{13}$$

Substituting (13) and (12) into (2) gives

$$\max_{\boldsymbol{B},\boldsymbol{G}} -\ln(\det \boldsymbol{B}) - \text{tr}\,[\boldsymbol{B}^{-1}\tilde{\boldsymbol{S}}_U] - \ln(\det \boldsymbol{G}) - \text{tr}\,[\boldsymbol{G}^{-1}\tilde{\boldsymbol{S}}_v] \tag{14}$$

where $\tilde{\boldsymbol{S}}_U = \boldsymbol{U}^H\boldsymbol{S}\boldsymbol{U}$ and $\tilde{\boldsymbol{S}}_v = \boldsymbol{V}^H\boldsymbol{S}\boldsymbol{V}$. Equation (14) represents two separate unconstrained maximizations. Both maximization are identical to maximizing likelihood when estimating an unstructured covariance matrix as discussed in [2] except that the transformed sample covariance matrices $\tilde{\boldsymbol{S}}_U$, $\tilde{\boldsymbol{S}}_V$ are present instead of $\boldsymbol{S}$. Thus, application of the results in [2] to this problem gives the ML estimates

$$\hat{\boldsymbol{B}} = \tilde{\boldsymbol{S}}_U$$
$$\hat{\boldsymbol{G}} = \tilde{\boldsymbol{S}}_V. \tag{15}$$

The signal and noise power estimates are given by

$$\hat{\sigma}_s^2 = \text{tr}\,\hat{\boldsymbol{B}} = \sum_{i=1}^{p} \boldsymbol{u}_i^H\boldsymbol{S}\boldsymbol{u}_i \tag{16a}$$

$$\hat{\sigma}_n^2 = \text{tr}\,\hat{\boldsymbol{G}} = \sum_{i=1}^{N-p} \boldsymbol{v}_i^H\boldsymbol{S}\boldsymbol{v}_i. \tag{16b}$$

Equations (16a) and (16b) are written in terms of projection operators as

$$\hat{\sigma}_s^2 = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{x}_m^H\boldsymbol{P}^s\boldsymbol{x}_m \tag{17a}$$

$$\hat{\sigma}_n^2 = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{x}_m^H\boldsymbol{P}^n\boldsymbol{x}_m \tag{17b}$$

where $\boldsymbol{P}^s = \boldsymbol{U}\boldsymbol{U}^H$ and $\boldsymbol{P}^n = \boldsymbol{V}\boldsymbol{V}^H$ are projection operators onto the signal and noise spaces. Note that as a result of the definitions of $\boldsymbol{V}$ and $\boldsymbol{U}$ we have $\boldsymbol{P}^n + \boldsymbol{P}^s = \boldsymbol{I}$ and therefore $\hat{\sigma}_s^2 + \hat{\sigma}_n^2 = \text{tr}\,\boldsymbol{S}$.

If the components within the signal and noise spaces are assumed to be uncorrelated, then the model for the data covariance is expressed as

$$\hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{V}^H \tag{18}$$

where $\boldsymbol{\Lambda} = \text{diag}\,(\lambda_1, \lambda_2, \cdots, \lambda_p)$ and $\boldsymbol{\Phi} = \text{diag}\,(\phi_1, \phi_2, \cdots, \phi_{N-p})$. Substituting (18) into (2) gives the maximum likelihood estimates

$$\hat{\lambda}_i = \boldsymbol{u}_i^H\boldsymbol{S}\boldsymbol{u}_i, \qquad 1 \le i \le p \tag{19a}$$

$$\hat{\phi}_i = \boldsymbol{v}_i^H \boldsymbol{S} \boldsymbol{v}_i, \qquad 1 \le i \le N - p \tag{19b}$$

$$\hat{\sigma}_s^2 = \sum_{i=1}^{p} \hat{\lambda}_i \tag{19c}$$

$$\hat{\sigma}_n^2 = \sum_{i=1}^{N-p} \hat{\phi}_i. \tag{19d}$$

Note that the estimates for σ_s^2 and σ_n^2 are identical to the general case in (16a), (16b) and the estimates for λ_i and ϕ_i correspond to the diagonal elements of $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{G}}$.

Example: Periodogram Analysis: As an example, we apply the results obtained here to the problem studied in [15], where $\hat{\boldsymbol{R}}$ is assumed circulant. $\boldsymbol{U}$ and $\boldsymbol{V}$ are composed of subsets of the DFT vectors corresponding to the frequencies in the signal and noise bands. Consider $M = 1$ with $\boldsymbol{x}_1$ composed of the time series samples $\boldsymbol{x}_1 = [x(0)x(1) \cdots x(N-1)]^T$. It is easy to verify that (19c), (19d) become

$$\hat{\sigma}_s^2 = \sum_{i=1}^{p} \frac{1}{N} \left| \sum_{k=0}^{N-1} x(k) e^{j2\pi k l_i / N} \right|^2 \tag{20a}$$

$$\hat{\sigma}_n^2 = \sum_{i=p+1}^{N} \frac{1}{N} \left| \sum_{k=0}^{N-1} x(k) e^{j2\pi k l_i / N} \right|^2 \tag{20b}$$

where the integers l_i, $1 \le i \le p$, and l_i, $p + 1 \le i \le N$, represent frequencies in the signal and noise bands, respectively. The signal and noise powers are the average of the periodogram over the appropriate frequencies. The $\hat{\lambda}_i$ and $\hat{\phi}_i$ represent estimates of the spectrum at index l_i and are given by the periodogram of the data.

Now consider the case where the covariance matrices corresponding to signal and noise components are known except for level. Represent $\boldsymbol{B} = \sigma_s^2 \boldsymbol{\Lambda}$ and $\boldsymbol{G} = \sigma_n^2 \boldsymbol{\Phi}$ with the diagonal matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ known. The data covariance is modeled by

$$\hat{\boldsymbol{R}} = \sigma_s^2 \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^H + \sigma_n^2 \boldsymbol{V} \boldsymbol{\Phi} \boldsymbol{V}^H. \tag{21}$$

This model corresponds to signal and noise components with known spectral shape but unknown power level. It is straightforward to show that the maximum likelihood estimates are given by

$$\hat{\sigma}_s^2 = \frac{1}{p} \operatorname{tr} \boldsymbol{\Lambda}^{-1} \boldsymbol{U}^H \boldsymbol{S} \boldsymbol{U} \tag{22a}$$

$$\hat{\sigma}_n^2 = \frac{1}{N - p} \operatorname{tr} \boldsymbol{\Phi}^{-1} \boldsymbol{V}^H \boldsymbol{S} \boldsymbol{V}. \tag{22b}$$

The $\boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{\Phi}^{-1}$ operations represent a whitening of the signal ($\boldsymbol{U}^H \boldsymbol{S} \boldsymbol{U}$) and noise ($\boldsymbol{V}^H \boldsymbol{S} \boldsymbol{V}$) components of the data.

Note that in the three models described by (11), (18), and (21), the signal and noise powers are estimated based on weighted combinations of $\boldsymbol{u}_i^H \boldsymbol{S} \boldsymbol{u}_i$ and $\boldsymbol{v}_i^H \boldsymbol{S} \boldsymbol{v}_i$. Furthermore, $\hat{\sigma}_s^2$ and $\hat{\sigma}_n^2$ are not dependent on each other as a result of the orthogonality of signal and noise space bases ($\boldsymbol{U}$ and $\boldsymbol{V}$).

B. Swept Frequency Cases

Apply a complex frequency shift to the model in (9). This is accomplished by multiplying the data by the matrix $\boldsymbol{E}(f) = N^{-1/2} \operatorname{diag}\{1, e^{j2\pi f}, e^{j4\pi f}, \cdots, e^{j2\pi f(N-1)}\}$. The covariance matrix (11) as a function of frequency becomes

$$\hat{\boldsymbol{R}}(f) = \boldsymbol{E}(f) \boldsymbol{U} \boldsymbol{B} \boldsymbol{U}^H \boldsymbol{E}^H(f) + \boldsymbol{E}(f) \boldsymbol{V} \boldsymbol{G} \boldsymbol{V}^H \boldsymbol{E}^H(f). \tag{23}$$

If $\boldsymbol{U}$ is a basis for signals which lie in the frequency band $-W \le f \le W$, then $\boldsymbol{E}(f_0)\boldsymbol{U}$ is a basis for signals which lie in the frequency band $-W + f_0 \le f \le W + f_0$. Estimating $\hat{\sigma}_s^2(f)$ as $\hat{\boldsymbol{R}}(f)$ is swept over $\{-1/2 \le f \le 1/2\}$ provides an estimate of the spectrum $\hat{\boldsymbol{S}}(f)$.

The maximum likelihood estimates derived earlier in this section utilize weighted combinations of $\boldsymbol{u}_i^H \boldsymbol{S} \boldsymbol{u}_i$ and $\boldsymbol{v}_i^H \boldsymbol{S} \boldsymbol{v}_i$. Since both $\boldsymbol{E}(f)\boldsymbol{U}$ and $\boldsymbol{E}(f)\boldsymbol{V}$ have orthonormal columns, estimates as a function of f are obtained by replacing $\boldsymbol{U}$ by $\boldsymbol{E}(f)\boldsymbol{U}$ and $\boldsymbol{V}$ by $\boldsymbol{E}(f)\boldsymbol{V}$. Define eigenspectra as

$$\left| y_i^u(f) \right|^2 = \boldsymbol{u}_i^H \boldsymbol{E}^H(f) \boldsymbol{S} \boldsymbol{E}(f) \boldsymbol{u}_i, \qquad 1 \le i \le p \tag{24a}$$

$$\left| y_i^v(f) \right|^2 = \boldsymbol{v}_i^H \boldsymbol{E}^H(f) \boldsymbol{S} \boldsymbol{E}(f) \boldsymbol{v}_i, \qquad 1 \le i \le N - p. \tag{24b}$$

Equation (16a) now becomes the spectrum estimate

$$\hat{S}(f) = \hat{\sigma}_s^2(f) = \sum_{i=1}^{p} \left| y_i^u(f) \right|^2. \tag{25}$$

In the model of (18) the eigenspectra correspond to $\hat{\lambda}_i$ in (19a) so the spectrum estimate is also given by (25). The spectrum estimate corresponding to (22a) is

$$\hat{S}(f) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i} \left| y_i^u(f) \right|^2. \tag{26}$$

In both cases the spectrum is estimated as a weighted combination of eigenspectra.

Consider the eigenspectra for $M = 1$ ($\boldsymbol{S} = \boldsymbol{x}\boldsymbol{x}^H$). Rewrite (24a) as

$$\left| y_i^u(f) \right|^2 = \left| \boldsymbol{u}_i^H \boldsymbol{E}^H(f) \boldsymbol{x} \right|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{x}(n) u_i(n) e^{-j2\pi f n}, \qquad 1 \le i \le p \tag{27}$$

where $u_i(n)$ represents the nth element of $\boldsymbol{u}_i^H$. The eigenspectra represent windowed periodograms of the data. As previously noted, the DPSS provide a basis for low-pass signals which are white on the band. Therefore, choosing the DPSS as a basis $\boldsymbol{U}$ results in Thomson's [1] eigenspectra. For $M > 1$, (24) corresponds to an average of M windowed periodograms.

The eigenspectra are also expressed as

$$\left| y_i^u(f) \right|^2 = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{x}_m^H \boldsymbol{P}_i(f) \boldsymbol{x}_m \tag{28}$$

where $\boldsymbol{P}_i(f)$ is a projection onto the space spanned by $\boldsymbol{E}(f)\boldsymbol{u}_i$. The norm of the data projected onto the space spanned by $\boldsymbol{E}(f)\boldsymbol{u}_i$ corresponds to the magnitude squared of the Fourier transform of the windowed data (see (27)). Thus, multiple window spectrum analysis is interpreted as defining a low rank model, projecting onto the components of this model and taking a weighted combination of the norm of the projected components as the spectrum estimate. A comparison of (8) and (26) indicates that the weightings $w_k(f)$ are equivalent to $(p\lambda_k)^{-1}$ in the structured covariance matrix model of (21).

C. Discussion

The advantage of low rank modeling is a trade of model bias for model variance [8]. Here model bias corresponds to error in representing the signal components in a band with a low rank model. Model variance corresponds to leakage of power from outside the band of interest to the estimate of in band power. Thomson [1] refers to model bias and variance as local and broad-band bias. Model bias results because $\boldsymbol{U}$ does not span all the components of a perfectly band-limited signal, unless $p = N$. Thus, some fraction of the power in the band is not represented. Model variance results because $\boldsymbol{U}$ is not perfectly band-limited and spans components which lie outside the band. As model rank (p) increases, model bias decreases but the model variance increases. The model variance also depends on the spectral level outside the band of interest and increases as the overall exterior spectrum level increases.

This suggests that the rank of the model should be chosen by considering the relative level of the spectrum within the band of interest to that outside the band of interest. When the internal spectrum level is large relative to the external level, the model variance will be small and a large rank model should be chosen to minimize model bias. When the internal spectrum level is small relative to the external level, the model variance is dominant and a small rank should be chosen. This rank reduction principle is evident in the adaptive weighting scheme of Thomson [1], as the higher order

eigenspectra are weighted less in regions where the spectrum is small.

IV. Center Frequency Estimation of a Signal With Known Spectral Shape

In this section, the swept frequency structured covariance matrix model is used to obtain the ML estimate of the center frequency of a signal with known spectral shape in white noise. As in the previous section, the estimate depends on a weighted combination of eigenspectra. The section concludes with an example of estimating the frequency of a complex exponential in white noise.

Assume that the covariance matrix of a rank p signal model in white noise for a center frequency of $f = 0$ is given by

$$\hat{\boldsymbol{R}}(0) = \boldsymbol{U}\,\boldsymbol{\Lambda}\,\boldsymbol{U}^H + \sigma_n \boldsymbol{I} \tag{29}$$

where $\boldsymbol{U}$ is N by p and $\boldsymbol{\Lambda}$ is p by p. Both are assumed known since they can be computed from the known spectral shape. Apply a complex frequency shift to obtain the covariance matrix as a function of center frequency.

$$\hat{\boldsymbol{R}}(f) = \boldsymbol{E}(f)\boldsymbol{U}\,\boldsymbol{\Lambda}\,\boldsymbol{U}^H\boldsymbol{E}^H(f) + \sigma_n \boldsymbol{I}. \tag{30}$$

A ML estimate of center frequency is achieved by estimating f in the model of (30).

Using (30) in (2) gives

$$\max_f \; -\ln\left|\hat{\boldsymbol{R}}(f)\right| - \operatorname{tr}\hat{\boldsymbol{R}}^{-1}(f)\boldsymbol{S} \tag{31}$$

for the ML estimate of f. The determinant of $\hat{\boldsymbol{R}}(f)$ is invariant to f so the maximization occurs only over the second term. Define $\boldsymbol{V}$ as before and write $\hat{\boldsymbol{R}}^{-1}(f)$ as

$$\begin{aligned}\hat{\boldsymbol{R}}^{-1}(f) = &\sum_{i=1}^{p} \frac{1}{\lambda_i + \sigma_n}\boldsymbol{E}(f)\boldsymbol{u}_i\boldsymbol{u}_i^H\boldsymbol{E}^H(f) \\ &+ \sum_{i=1}^{N-p}\frac{1}{\sigma_n}\boldsymbol{E}(f)\boldsymbol{v}_i\boldsymbol{v}_i^H\boldsymbol{E}^H(f).\end{aligned} \tag{32}$$

Substituting (32) into (31) and simplifying yields

$$\min_f \sum_{i=1}^{p}\frac{1}{\lambda_i + \sigma_n}\left|y_i^u(f)\right|^2 + \frac{1}{\sigma_n}\sum_{i=1}^{N-p}\left|y_i^v(f)\right|^2 \tag{33}$$

where $|y_i^u(f)|^2$ and $|y_i^v(f)|^2$ are eigenspectra as defined in (24). It is straightforward to show that

$$\sum_{i=1}^{N-p}\left|y_i^v(f)\right|^2 = \operatorname{tr}\boldsymbol{S} - \sum_{i=1}^{p}\left|y_i^u(f)\right|^2. \tag{34}$$

Substituting (34) into (33) and simplifying we obtain

$$\max_f \sum_{i=1}^{p}\left[\frac{\lambda_i}{\lambda_i + \sigma_n}\right]\left|y_i^u(f)\right|^2. \tag{35}$$

Thus, the ML estimate of center frequency is given by the frequency at which a weighted combination of eigenspectra is maximum. The weighting is proportional to the ratio of signal-to-signal plus noise power in each component of $\boldsymbol{U}$. If the noise power is very small relative to the λ_i, the eigenspectra are weighted uniformly, but if σ_n is very large relative to λ_i, the eigenspectra are weighted proportionally to the signal eigenvalues (λ_i).

Example: Complex Exponential in Noise: The covariance matrix for a complex exponential located at $f = 0$ in white noise is given by

$$\hat{\boldsymbol{R}}(0) = \sigma_s^2\mathbf{1}\mathbf{1}^H + \sigma_n\boldsymbol{I} \tag{36}$$

where $\mathbf{1}$ is a vector with all unity elements. In this case $p = 1$ and $\boldsymbol{U} = N^{-1/2}\mathbf{1}$. Equation (35) becomes

$$\max_f \left|y_1^u(f)\right|^2. \tag{37}$$

For $M = 1$

$$\left|y_1^u(f)\right|^2 = \frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)e^{-j2\pi nf}\right|^2. \tag{38}$$

This corresponds to the well-known result that the ML estimate of the frequency of a complex exponential in white noise is given by the frequency at which the periodogram is maximum [16].

V. Summary

This paper presents an approach to spectrum analysis based on maximum likelihood estimation of a structured covariance matrix. The structured covariance matrix is a low rank approximation of the covariance matrix for a narrow-band signal in noise, and the power in narrow bands which are swept in frequency provide the spectrum estimate. The ML estimate of the power in each band is obtained by projecting the data onto the components of the low rank model and taking a weighted combination of the power in each model component. This is equivalent to estimating the spectrum as a weighted combination of eigenspectra and is intimately related to the multiple window spectrum analysis techniques of Thomson [1]. The maximum likelihood estimate of center frequency for a low rank signal with known spectral shape in white noise is also given by a weighted combination of eigenspectra. The use of low rank models permits the trading of modeling error within the band for leakage from outside the band of interest.

Acknowledgment

The authors thank Prof. D. Tufts for suggesting the center frequency estimation problem and one of the reviewers for pointing out an error in an earlier version of the manuscript.

References

[1] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055-1096, Sept. 1982.

[2] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, pp. 963-974, Sept. 1982.

[3] R. D. Bock, "Components of variance analysis as a structural and discriminal analysis of psychological tests," *Brit. J. Statist. Psych.*, vol. 13, pp. 151-163, 1960.

[4] R. D. Bock and R. E. Bargmann, "Analysis of covariance structures," *Psychometrika*, vol. 31, pp. 507-534, Dec. 1966.

[5] T. W. Anderson, "Statistical inference for covariance matrices with linear structure," *Multivariate Analysis II*, P. R. Krishnaiah, Ed. New York: Academic, 1969, pp. 55-66.

[6] T. W. Anderson, "Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices," in *Essays in Probability and Statistics*, R. C. Bose, Ed. Chapel Hill, NC: University of North Carolina Press, 1970, pp. 1-24.

[7] T. W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure," *Ann. Statist.*, vol. 1, pp. 135-141, 1973.

[8] L. L. Scharf and D. W. Tufts, "Rank reduction for modeling stationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 350-355, Mar. 1987.

[9] H. Wang and G. H. Wakefield, "Signal subspace approximation for line spectrum estimation," in *Proc. ICASSP 87*, Apr. 1987, pp. 2054-2057.

[10] L. L. Scharf and P. J. Tourtier, "Maximum likelihood identification of structured correlation matrices for spectrum analysis," in *Proc. 20th Asilomar Conf. Signals, Syst., Comput.*, Nov. 1986.

[11] P. J. Tourtier and L. L. Scharf, "Maximum likelihood identification of correlation matrices for estimation of power spectra at arbitrary resolutions," in *Proc. ICASSP 87*, Apr. 1987, pp. 2066-2069.

[12] L. L. Scharf and B. D. Van Veen, "Low rank detectors for Gaussian random vectors," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-35, pp. 1579-1582, Nov. 1987.

[13] N. Q. A, "On the uniqueness of the maximum likelihood estimate of structured covariance matrices," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-32, pp. 1249-1251, Dec. 1984.

[14] D. Slepian, "Prolate spheriodal wave functions, Fourier analysis, and uncertainty-V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, pp. 1371-1430, May 1978.

[15] A. Dembo, "The relation between maximum likelihood estimation of structured covariance matrices and periodograms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1661-1662, Dec. 1986.

[16] S. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[17] C. T. Mullis and L. L. Scharf, "Quadratic estimators of the power spectrum," Dep. Elec. Comput. Eng., Univ. of Colorado, CO, Tech. Rep., Apr. 1989.

A New Bit Reversal Algorithm

JAMES S. WALKER

***Abstract*—This correspondence describes a new bit reversal permutation algorithm. Such algorithms are needed for radix 2 (or radix B) fast Fourier transforms (FFT) or fast Hartley transforms (FHT). This algorithm is an alternative to one described by Evans. A BASIC version of this algorithm ran slightly faster than the BASIC version of Evans' algorithm given in Bracewell's book, with some time savings for odd powers of 2. This new algorithm also allows for precomputation of seed tables up to one higher power of 2 than Evans' algorithm.**

I. Introduction

We shall concentrate on radix 2, since radix B involves no great changes and radix 2 is very widely used for fast Fourier transforms (FFT's) and fast Hartley transforms (FHT's). In almost all FFT's and FHT's, either the input or output is permuted according to the reverse ordering of the binary expansions of the data indices. If a data index m has the binary expansion $m = z_1 z_2 \cdots z_{R-1} z_R$ (base 2) where each z_i is 0 or 1 and $N = 2^R$ (N being the number of elements in the data), then the digits (bits) of m are reversed, obtaining $P_N(m)$ where $P_N(m) = z_R z_{R-1} \cdots z_2 z_1$ (base 2). The data having index m is then swapped with the data having index $P_N(m)$.

We will now describe an efficient means of performing this permutation. Suppose $N = 2^R$ where R is even (the case of R odd will be discussed later). Then a typical index m will have a binary expansion $m = x_k \cdots x_1 y_k \cdots y_1$ (base 2) where $k = R/2$. Then $P_N(m)$ is obtained by *drawing a mirror* through the middle of its bits and doing a reflection (the first arrow indicates the mirror position)

$$m = x_k x_{k-1} \cdots x_1 \uparrow y_k k_{k-1} \cdots y_1$$
$$\rightarrow y_1 \cdots y_{k-1} y_k x_1 \cdots x_{k-1} x_k = P_N(m).$$

In ordinary notation this looks like ($S = N^{1/2}$)

$$m = L \cdot S + M \rightarrow P_N(m) = P_S(M) \cdot S + P_S(L)$$
$$[0 \le L \le S - 1, 0 \le M \le S - 1]. \qquad (1)$$

This factors P_N into two bit reversals P_S where $S = N^{1/2}$.

Our discussion, up to this point, is the same as the initial discussion in [1, sec. II]. *Evans' program*, see especially [2, pp. 119–120, lines 9412–9429], *is an implementation of* (1).

We will now discuss an alternative to programming (1). To describe this alternative, we will use the notation $m \leftrightarrow n$ to denote a pair of indices whose data, f_m and f_n, are to be swapped. For convenience, we will simply refer to $m \leftrightarrow n$ as a swap, even though f_m and f_n are actually swapped.

The alternative to programming (1) arises by *parameterizing the set of swaps*, $\{m \leftrightarrow n\}$. The key to this is the *reflexive* numbers, those numbers for which $m = P_N(m)$. Clearly, those numbers have the form $m = L \cdot S + P_S(L)$ [$0 \le L \le S - 1$]. There are precisely $N^{1/2} = S$ of these reflexive numbers, which, if we label them by $[L, P_S(L)]$, *lie along the main diagonal* of the pairs in Table I.

Manuscript received November 7, 1988; revised September 5, 1989.
The author is with the Mathematics Department, University of Wisconsin-Eau Claire, Eau Claire, WI 54702-4004.
IEEE Log Number 9036517.

TABLE I
PARAMETERIZATION OF SWAPS

L \ K	0	1	....	K	....	S-1
0	[0,P_S(0)]	[0,P_S(1)]	....	[0,P_S(K)]	....	[0,P_S(S-1)]
1	[1,P_S(0)]	[1,P_S(1)]	....	[1,P_S(K)]	....	[1,P_S(S-1)]
2	[2,P_S(0)]	[2,P_S(1)]	....	[2,P_S(K)]	....	[2,P_S(S-1)]
⋮		.....				
L	[L,P_S(0)]	[L,P_S(1)]	....	[L,P_S(K)]	....	[L,P_S(S-1)]
⋮		.....				
S-1	[S-1,P_S(0)]	.....		[S-1,P_S(K)]	...	[S-1,P_S(S-1)]

Each element $[L, P_S(K)]$ in Table I determines a swap $m \leftrightarrow n$ defined by

$$m = L \cdot S + P_S(K) \leftrightarrow n = K \cdot S + P_S(L)$$
$$[0 \le L \le S - 1, 0 \le K \le S - 1]. \qquad (2)$$

The upper triangular portion of Table I [$1 \le K \le S - 1, 0 \le L \le K - 1$] determines the same set of swaps $\{m \leftrightarrow n\}$ as the lower triangular portion [$1 \le L \le S - 1, 0 \le K \le L - 1$]. While, as stated above, the main diagonal of Table I [$0 \le L = K \le S - 1$] determines the unnecessary swaps $n \leftrightarrow n$. Due to this symmetry, and the fact that all swaps in P_N are determined by (2), it follows that the pairs

$$m = L \cdot S + P_S(K) \leftrightarrow n = K \cdot S + P_S(L)$$
$$[1 \le L \le S - 1, 0 \le K \le L - 1] \qquad (3)$$

describe all the swaps that occur in P_N. Moreover, since there are $S(S-1)/2 = (N - N^{1/2})/2$ swaps in P_N (ignore the $S = N^{1/2}$ reflexive numbers and group the remaining $N - N^{1/2}$ elements in pairs) *and* the same number of swaps is given in (3), it also follows that (3) describes all the swaps in P_N in a *one to one* fashion.

For $N = 2^R$, where R is odd, the mirror would fall on the central bit in the number (the arrows indicate the mirror positions)

$$x_k \cdots x_1 \overset{0}{\uparrow} y_k \cdots y_1 \quad \text{or} \quad x_k \cdots x_1 \overset{1}{\uparrow} y_k \cdots y_1$$

and then we obtain ($S = N^{1/2}$, $W = 2S$)

$$m = L \cdot W + P_S(K) \leftrightarrow n = K \cdot W + P_S(L)$$
$$\text{and} \quad m + S \leftrightarrow n + S$$
$$[1 \le L \le S - 1, 0 \le K \le L - 1] \qquad (3')$$

where the first swap in (3') occurs when 0 is the central bit and the second swap occurs when 1 is the central bit.

II. Programming and Timing

Programming of (3) and (3') is quite simple. In Table II there is a compiled BASIC program FPW which was used in timing experiments. The permutation $P_S(\ )$ is called $J(\ )$ in FPW. The computation of $J(\ )$ is a version of a program of Buneman (see [3]) which was given to me by Bracewell [4]. The variable names $M7$ (also $M6$) and $N6$ correspond to m and n in (3) and (3'). Variable names $M5$ and $N5$ correspond to $m + S$ and $n + S$ in (3'), while $N9$ is used for S in (3) and (3') and $N8$ is used for W in (3'). Programming was done so that multiple additions are performed instead of (often more time-consuming) multiplications. A precompiled version of FPW was timed against a precompiled BASIC version of Evans' algorithm (see [2, pp. 119–120, lines 9412–9429]