ADAPTIVE DATA ORTHOGONALIZATION

N. L. Owsley

Naval Underwater Systems Center
New London, Conn. 06320

ABSTRACT

The decomposition of vector time series data into orthogonal components can be applied in both temporal and spatial discrete frequency analysis. If the observed multidimensional data is non-stationary, then adaptive procedures can be used for estimation of the eigendata. This paper presents the relationship between multicomponent, spectral signals in noise and the corresponding eigendata. Two adaptive realizations of the eigendata estimation process are considered. Examples are given which allow a comparison between signal detection by data orthogonalization, power spectrum estimation and two channel magnitude squared coherence computation.

I INTRODUCTION

The orthogonal decomposition of vector time series data is an objective of specific signal processing endeavors in both temporal frequency and spatial wavenumber analysis. Frequency diverse spectral set extraction(1) and antenna array processing for ambient noise analysis(2) and source location (3) are examples of such applications. In addition, the general field of pattern recognition technology frequently utilizes data orthogonalization for feature extraction(4). If the observed vector processes are statistically non-stationary, then adaptive schemes for estimating the underlying characteristic roots and vectors of this data are required. The literature is abundant on both iterative and transformation procedures for the computation of the eigenvalues and eigenvectors of general matrices(5). This paper considers an adaptive algorithm for the direct estimation of the characteristic roots and vectors, hereafter referred to as eigendata, directly from the sampled data without an intermediate data correlation matrix estimation step as would be required with either an iterative or transformation scheme. Specifically, the eigendata corresponding to a small number of the largest eigenvalues shall be estimated using a stochastic gradient ascent algorithm(3) which operates directly on the sampled data and uses a minimum of computation and data storage. This approach is compared with an orthogonal iteration procedure which requires sufficient data storage and computation rate to provide a data correlation matrix estimate. It is shown that the eigendata estimation scheme using gradient ascent, while characterized by simplicity of implementation, suffers from relatively slow convergence at low SNR and greater estimate mean square error (MSE) than the iterative estimator. It is suggested that a combination scheme using the gradient ascent estimators as initial vectors for an orthogonal iteration could be an effective implementation.

II ORTHOGONALIZATION OF MULTIPLE COMPONENT SIGNAL DATA (1)

Consider the N-dimensional, zero mean complex random vector process $\underline{X}(n)$ sampled at time n. Let the process be stationary and characterized by the correlation matrix

$$R = E\{\underline{X}(n)\underline{X}(n)^{\dagger}\} \tag{1}$$

where $E\{\cdot\}$ and $\dagger$ indicate the statistical expectation and matrix complex conjugate transpose operators respectively. The matrix R can be expressed as

$$R = \sum_{i=1}^{N} \lambda_i \underline{M}_i \underline{M}_i^{\dagger} \tag{2}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots . \geq \lambda_N$ are the eigenvalues and $\underline{M}_1, \underline{M}_2, \ldots, \underline{M}_N$ the corresponding orthonormal eigenvectors of R. Let the sample vector process be of the form

$$\underline{X}(n) = \sum_{i=1}^{K} s_i(n)\underline{E}_i + \underline{N}(n) \tag{3}$$

were $s_i(n)$ is a zero mean, complex random variable, $\{\underline{E}_i\}_{i=1}^{K}$ is an orthonormal vector set spanning a $K \leq N$ dimensional, complex vector (signal) subspace and $\underline{N}(n)$ is a sample vector from an identically distributed and statistically independent vector gaussian random process with zero mean and variance σ_o^2. If the crosscorrelation $p_{ij} = E\{s_i(n)s_j(n)^{\dagger}\}$ between the ith and jth signal coefficients is introduced, then the correlation matrix for the process described in (3) becomes

CH1285-6/78/0000-0109$00.75©1978IEEE

$$R = \sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij}\underline{E}_i\underline{E}_j^{\dagger} + \sigma_o^2 I_N \qquad (4)$$

where I_N is an N-dimensional identity matrix. Now let the signal coefficients be characterized by coherent sets according to the correlation coefficient values

$$\frac{|p_{ij}|^2}{S_i S_j} = \begin{cases} 1 \text{ for } i \text{ and } j \in \Omega_l \\ 0 \text{ for } i \in \Omega_l \text{ and } j \in \bar{\Omega}_l \end{cases} \qquad (5)$$

In (5) $p_{ii} = S_i$ and the notation $i(j)\in\Omega_l(\bar{\Omega}_l)$ implies the ith(jth) signal coefficient $s_i(n)$ ($s_j(n)$) is (is not) a member of a coherent subset $\{s_i(n)/i\in\Omega_l\}$. With p_{ij} of the form $p_{ij}=[S_iS_j]^{\frac{1}{2}}$ and the stipulation that L coherent signal subsets are extant, the following expressions for the matrix R eigendata can be obtained

$$\lambda_l = \sum_{i\in\Omega_l} S_i + \sigma_o^2$$

$$\underline{M}_l = \frac{1}{[\sum_{i\in\Omega_l} S_i]^{\frac{1}{2}}} \sum_{i\in\Omega_l} S_i^{\frac{1}{2}}\underline{E}_i \qquad (6)$$

for $1\le l\le L$. For $L<l\le N$ there results $\lambda_l=\sigma_o^2$ with the corresponding $\underline{M}_l$ an unspecified vector in an orthonormal set $\{\underline{M}_i\}_{i=L+1}^{N}$ defining a vector space orthogonal to the space spanned by $\{\underline{M}_i\}_{i=1}^{L}$.

Examining (6) it is seen that all the information concerning the ith signal coefficient power S_i, component vector $\underline{E}_i$ and coherent subset membership $i\in\Omega_l$ can be extracted from the first $L \le K$ eigendata pairs. It therefore follows that an analysis of $\underline{X}(n)$ need be concerned with estimation of the first L characteristic roots and vectors of the correlation matrix R to obtain these significant signal component descriptors.

III ADAPTIVE REALIZATIONS

A constrained gradient ascent algorithm for estimation of $\underline{M}_l(n)$ at time n can be specified according to the objective that an estimate $\underline{W}_l(n)$ of $\underline{M}_l(n)$ be found such that for R(n) an estimate of R at time n the real scalar

$$\lambda_l(n) = \underline{W}_l(n)^{\dagger}R(n)\underline{W}_l(n) \qquad (7)$$

be maximized with respect to $\underline{W}_l(n)$ subject to the l-1 orthogonality constraints

$$\underline{W}_l(n)^{\dagger}\underline{M}_i(n) = 0 \quad (i = 1,2,\ldots,l-1) \qquad (8)$$

and the normality constraint

$$|\underline{W}_l(n)|^2 = 1. \qquad (9)$$

An algorithm to accomplish this objective is

$$\left[\underline{W}_l(n)=P_l(n)[\underline{W}_l(n-1)+\mu R(n)\underline{W}_l(n-1)]\right]_{|\underline{W}_l(n)|^2=1} \qquad (10)$$

where

$$P_l(n) = I_N - \sum_{i=1}^{l-1} \underline{M}_i(n)\underline{M}_i(n)^{\dagger}, \qquad (11)$$

and μ is a positive constant which controls the rate of convergence and the MSE of the steady state estimate. Equation (10) is justified by noting that the gradient of $\lambda_l(n)$ with respect to $\underline{W}_l(n)$ is proportional to $R(n)\underline{W}_l(n)$. Thus, adjusting $\underline{W}_l(n-1)$ by an increment $\mu\, R(n)\underline{W}_l(n-1)$ will yield an increased value of the estimated eigenvalue $\lambda_l(n)$. However, this updated eigenvector estimate must remain orthogonal to the eigenvectors $\{\underline{M}_i(n)\}_{i=1}^{l-1}$ and have unit magnitude. It can be verified that premultiplication of the unconstrained, updated vector by the projection matrix $P_l(n)$ and a straight-forward normalization accomplish both of the required constraints. In an actual implementation the noisy estimate of the correlation matrix $R(n)=\underline{X}(n)\underline{X}(n)^{\dagger}$ is used with attendant savings in computational load compared to a realization which first estimates R(n). In addition, the eigenvector estimates $\underline{M}_i(n)=\underline{W}_i(n)$, i=1,2,..., l are used in the computation of (10).

In order to discuss the convergence of the mean of (10) it is first necessary to introduce some matrix definitions. Specifically, let λ be an N-by-N diagonal matrix with λ_i as the ith diagonal element. Also, let M be an N-by-N matrix with $\underline{M}_i$ as the ith column. Finally, let C_l be an N-by-(l-1) matrix

$$C_l = \begin{bmatrix} I_{l-1} \\ \hdashline 0 \end{bmatrix}. \qquad (12)$$

Assuming that the first l-1 eigenvectors are known and that $\underline{W}_l(n-1)$ and $\underline{X}(n)$ are uncorrelated allows the expectation of (10) exclusive of the normality constraint to be written

$$E\{\underline{W}_l(n)\} = \left[I_N - MC_lC_l^{\dagger}M^{\dagger}\right]\left[I_N+\mu\, M\lambda M^{\dagger}\right]E\{\underline{W}_l(n-1)\}. \qquad (13)$$

Letting $\underline{W}_l(0)=M\underline{A}(0)$ with all elements in $\underline{A}(0)$ real, (13) can be iterated n times to yield

$$E\{\underline{W}_l(n)\} = M\left[I_N-C_lC_l^{\dagger}\right]\left[I_N+\mu\lambda\right]^n \underline{A}(0). \qquad (14)$$

For distinct eigenvalues and n sufficiently large

$$E\{\underline{W}_l(n)\} \cong (1+\mu\,\lambda_l)^n a_{l0}\underline{M}_l \qquad (15)$$

where a_{l0} is the lth element in $\underline{A}(0)$. The normalized value of the right hand side of (15) is readily seen to be $\underline{M}_l$. Examination of (14) indicates that the convergence of the lth eigenvector component is exponential with resolution from the mth component occuring at a rate controlled by the ratio $(1+\mu\ \lambda_1)/(1+\mu\ \lambda_m)$. Thus, with either λ_1 and λ_m ($1<m$) small or nearly equal and μ sufficiently small, it is evident that convergence could be quite slow.

In view of the potential convergence rate problems of the gradient eigenvector estimation algorithm, the orthogonal iteration algorithm

$$\left[\underline{W}_1(k)=P_1(k)R(n)\underline{W}_1(k-1)\right]_{|\underline{W}_1(k)|^2=1} \tag{16}$$

is considered. This algorithm is applied periodically to an estimate of the correlation matrix R, namely R(n), which, for example, could be computed as

$$R(n) = (1-\mu)R(n-1)+ \mu\ \underline{X}(n)\underline{X}(n)^\dagger \qquad (0<\mu<1). \tag{17}$$

In this approach n is still the time sample index, however, k is an iteration index for application of (16) at sample points separated by $n_d=1/\ln(1/1-\mu)$ samples equivalent to a single time constant. The convergence of (16) at time n can be examined as above in terms of the expectation of the kth unnormalized iterate. Assuming $\underline{M}_i$, $i=1,2,\ldots,l-1$ have already been determined, it is straight-forward to show that after k iterations

$$E\{\underline{W}_l(k)\} = M[I_N-C_lC_l^\dagger]\ \lambda^k\underline{A}(0). \tag{18}$$

Assuming that $E\{R(n)\} = R = M\lambda M^\dagger$ has distinct roots $\lambda_1>\lambda_2>\ldots>\lambda_N$ it is seen that for $k>>1$

$$E\{\underline{W}_l(k)\} \cong a_{l0}\lambda_l^k\underline{M}_l \tag{19}$$

with normalized value $\underline{M}_l$. It is important to note that (18) approaches $\underline{M}_l$ to within a scalar multiple independent of the sample time n, but rather depends on k which can be made as large as required to obtain a satisfactory accuracy. Furthermore, the convergence rate at time n is again exponential, but with resolution of the lth component from the mth component occurring at a rate controlled by the ratio λ_l/λ_m instead of $(1+\mu\ \lambda_l)/(1+\mu\ \lambda_m)$. This allows for convergence of the orthogonal iteration algorithm even for small eigenvalues and long averaging times (small μ values) provided the eigenvalues are distinct.

The above treatment of convergence of the mean for the eigenvector estimation algorithms is oversimplified with respect to the application of the normalization constraint. A rigorous treatment of the convergence of (10) involves consideration of a nonlinear algorithm due to the quadratic form of the normalization constraint. However, exposure of the convergence rate and eigenvector resolution mechanisms afforded by this mean value consideration is thought to be valid. With respect to steady state MSE for the eigendata estimates some qualitative statements are offered which compare the performance of the stochastic gradient and orthogonal iteration approaches. Specifically, it is observed that the steady state MSE for a noisy gradient type algorithm estimate is proportional to N, the dimension of the sampled process, and to μ, the gradient step size control parameter[6].

In contrast, assuming distinct eigenvalues, the orthogonal iteration estimate by making the number of iterations k arbitrarily large can be made to have a MSE a function only of μ which determines the time constant used in computing the estimate R(n) of R (17). Thus for the same effective averaging time the eigendata estimates obtained using orthogonal iteration should have less steady state MSE than the corresponding estimate using constrained gradient ascent where MSE depends on N. These statements are experimentally supportable as indicated in the following section.

IV EXAMPLE : COMPLEX SPECTRUM PROCESSING

In this section the spectrum analysis system illustrated in Figure 1 is used to evaluate data orthogonalization as a signal detection processor. Two channels consisting of an identical signal with four specular components, but with uncorrelated identically distributed, zero mean noises additive in each channel are simulated with eight bits of precision. The simulated time domain signal plus noise is fast Fourier transformed (FFT) in segments with the output vectors from the FFT serving as the sampled data vector $\underline{X}(n)$ where n is the FFT time segment index. Three processors are compared. With $X_i(f,n)$ the fth term in the ith channel FFT output vector for time segment n, the detector outputs are given by
Power Spectrum (PS) :

$$P(f,n) = (1-\mu)P(f,n-1) + \mu\ |X_1(f,n)|^2 \tag{20}$$

Orthogonalization (gradient ascent-OG, iteration -OI):

$$O(f,n,l) = |W_1(f,n)|^2 \quad (l=1,2) \tag{21}$$

Magnitude squared coherence (CP)[7]:

$$C(f,n)=|C_{12}(f,n)|^2/C_{11}^2(f,n)C_{22}^2(f,n) \tag{22}$$

where

$$C_{ij}(f,n)=(1-\mu)C_{ij}(f,n-1)+\mu\ X_i(f,n)X_j^\dagger(f,n) \tag{23}$$

An example of the various detector outputs is given in Figure 2. Four spectral components with the lowest three in the same spectral set are illustrated. OG1 and OI1 indicate the detector outputs for the first eigenvector and OG2 and OI2 for the second. The FFT resolution was 1 Hz with signal-to-noise ratios (dB/Hz) of the four components of 3, 3, 3 and 1.5 from lowest to highest frequency respectively. Several aspects of data orthogonalization have been substantiated experimentally. Specifically, the following predicted features of data orthogonalization detector performance are verified experimentally.

- coherent signal components appear in the same eigenvector
- the orthogonalization detector behaves as a correlation processor where coherent components are cross-correlated as in a magnitude squared coherence processor
- data orthogonalization by orthogonal iteration is superior to orthogonalization by gradient ascent with respect to convergence rate and MSE

The computational load of the gradient scheme for adaptive data orthogonalization is almost negligible compared to an iterative realization. However, the performance of the iterative scheme is significantly better and therefore the iterative approach is preferred for the extraction of low SNR signal components. The number of iterations required for the OI detector realization can be reduced by using the OG eigenvectors as initial vectors for the OI processor iteration.

V CONCLUSION

Two realizations of an adaptive data orthogonalization scheme for vector time series have been considered. On the basis of convergence rate and MSE the method of orthogonal iteration is preferred to the gradient ascent realization in spite of significantly greater computational requirements. A data orthogonalization processor used as a detector performs as a coherence detector. A magnitude squared coherence detector derives a noisy coherent reference for correlation from the existance of an additional channel with the same signal. The orthogonalization processor derives this noisy reference from coherent signal components within a single channel.

VI ACKNOWLEDGEMENT

The author is grateful to the Norwegian Defence Research Establishment (NDRE), Horten, Norway for partial support of the work presented herein. The author was assigned to NDRE in the period 1976-77.

REFERENCES

(1) N L Owsley, "Spectral Set Extraction", Aspects of Signal Processing, Edited by G Tacconi, D Reidel Publ, Vol 2, pp 469-475, 1977

(2) N C Yen, "Ambient Sea Noise Directionality": Measurement and Processing. Naval Underwater Systems Center, New London, Conn, Tech Rpt 554 28 February 1977

(3) N L Owsley, "A Recent Trend in Adaptive Spatial Processing for Arrays : Constrained Adaptation", Signal Processing, Edited by J W R Griffiths et al, Academic Press, pp 605-618, 1973

(4) L N Kanal, "Interactive Pattern Analysis and Classification system. A survey and Commentary". Proc of IEEE, Vol 60, No 10, pp 1200-1216. October 1972

(5) I S Duff, "A Survey of Sparse Matrix Research", Proc of IEEE, Vol 65, No 4, pp 500-536, Apr 1977

(6) B Widrow et al, "Adaptive Antenna Systems", Proc of IEEE, Vol 55, No 12, pp 2143-2158, December 1967

(7) G C Carter, "Receiver Operating Characteristics for a Linearly Thresholded Coherence Estimation Detector". IEEE Trans on ASSP, pp 90-93. February 1977

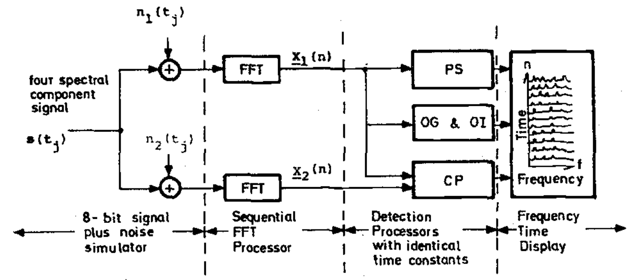

Figure 1 Complex spectrum processing by Power Spectrum (PS) estimation, Orthogonalization by Gradient (OG) ascent, Orthogonalization by Iteration (OI) and magnitude squared Coherence Processing (CP).

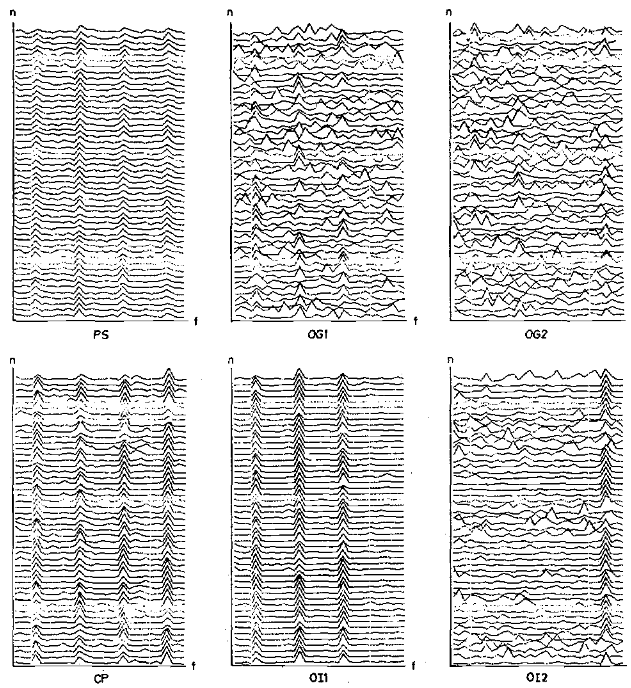

Figure 2 Frequency (f) versus time (n) outputs for PS, OG1 and OI1 (eigenvector for largest eigenvalue λ_1), OG2 and OI2 (eigenvector for second largest eigenvalue λ_2), and CP systems.