Maximizing (17) over P_i is equivalent to choosing P_i to be equal to the minimum of (18), divided by N_i.

If $N_i > i$ for all i, so that $N_1 \geq N_2 \geq \cdots \geq N_M > M$ (i.e., every component is observed at least $M + 1$ times), and if the true covariance matrix is positive definite, then, with probability one, the P_i are strictly positive (so the estimated covariance matrix is positive definite).

The decoupling of b_i, P_i, and the ith row of L, for different i, facilitates the computation of Cramer-Rao bounds, since it imparts a block-diagonal structure to the Fisher information matrix.

V. Example: ML Estimates for Bivariate, Zero-Mean Case

It is instructive to consider in detail the bivariate case (i.e., $M = 2$). We further simplify the problem by assuming that the true mean is zero (i.e., $\boldsymbol{a} = \boldsymbol{o}$). Here we present the exact ML estimates for the elements of the covariance matrix.

For $M = 2$, and with the assumption that $\boldsymbol{b} = \boldsymbol{o}$, it is straightforward to obtain the ML estimates for L and P. Then transforming back to K gives the following exact ML estimates:

$$\hat{K}_{11} = \left[\frac{1}{N_1} \cdot \sum_{n=1}^{N_1} x_{1n}^2\right] \tag{19}$$

$$\hat{K}_{12} = \left[\frac{1}{N_2} \cdot \sum_{n=1}^{N_2} x_{1n} \cdot x_{2n}\right] \cdot \frac{\left[\frac{1}{N_1} \sum_{n=1}^{N_1} x_{1n}^2\right]}{\left[\frac{1}{N_2} \sum_{n=1}^{N_2} x_{1n}^2\right]} \tag{20}$$

$$\begin{aligned}\hat{K}_{22} = &\left[\frac{1}{N_2} \cdot \sum_{n=1}^{N_2} x_{2n}^2\right] \\ &- \frac{\left[\frac{1}{N_2} \sum_{n=1}^{N_2} x_{1n} \cdot x_{2n}\right]^2}{\left[\frac{1}{N_2} \sum_{n=1}^{N_2} x_{1n}^2\right]^2} \cdot \left\{\left[\frac{1}{N_2} \sum_{n=1}^{N_2} x_{1n}^2\right]\right. \\ &\left. - \left[\frac{1}{N_1} \cdot \sum_{n=1}^{N_1} x_{1n}^2\right]\right\}.\end{aligned} \tag{21}$$

It is apparent that, for $N_1 > N_2$, the ML estimates for K_{12} and K_{22} are exceedingly complicated. In contrast, the ML estimates for L and P are relatively simple. Note that the ML estimates for K_{12} and K_{22} involve all of the observations.

VI. Summary and Conclusions

We have obtained the exact maximum likelihood estimates for the mean vector and the covariance matrix for a class of problems where not every component of each realization of the random vector is observed. In contrast to ad hoc estimates, the ML estimate for the covariance matrix is guaranteed to be positive definite.

It would be reasonable to apply our algorithm to non-Gaussian data as well as to Gaussian-distributed data. The estimator would not be maximum-likelihood for non-Gaussian data, but it would produce consistent estimates. Error norms other than least squares could be used to estimate the regression coefficients.

Our algorithm is applicable when it is possible to order the components of the random vector such that the set of realizations for which the ith component is available is a subset of the set of realizations for which the $(i - 1)$th component is available, for $2 \leq i \leq M$. Such cases appear in practice when multiple sensors are used sequentially, with each sensor observing a subset of the objects that were observed by the previous sensor. Our algorithm is not applicable to the data set shown in Fig. 2 because there is no simple expression for the likelihood function in terms of the Cholesky factors. Here an iterative ascent algorithm such as EM could still be used.

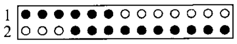

Fig. 2. Structure of data set comprising independent realizations of a bivariate random vector; some realizations are missing the first component and some realizations are missing the second component; there is no simple expression for the likelihood function in terms of the Cholesky factors.

References

[1] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.

[2] J. Burg, D. Luenberger, and D. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, Sept. 1982.

[3] B. Van Veen and L. Scharf, "Estimation of structured covariance matrices and multiple window spectrum analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 8, pp. 1467-1472, Aug. 1990.

Signal Detection via Spectral Theory of Large Dimensional Random Matrices

Jack W. Silverstein and Patrick L. Combettes

***Abstract*—Results on the spectral behavior of random matrices as the dimension increases are applied to the problem of detecting the number of sources impinging on an array of sensors. A common strategy to solve this problem is to estimate the multiplicity of the smallest eigenvalue of the spatial covariance matrix R of the sensed data from the sample covariance matrix $\hat{R}$. Existing approaches rely on the closeness of the noise eigenvalues of $\hat{R}$ to each other and, therefore, the sample size has to be quite large when the number of sources is large in order to obtain a good estimate. The theoretical analysis presented in this correspondence focuses on the splitting of the spectrum of $\hat{R}$ into noise and signal eigenvalues. It is shown that, when the number of sensors is large, the number of signals can be estimated with a sample size considerably less than that required by previous approaches.**

I. Introduction

In many signal processing applications, a fundamental problem is the determination of the number of signals impinging on an array

Manuscript received May 23, 1990; revised July 8, 1991. The work of the first author was supported by the National Science Foundation under Contract DMS-8903072. The simulations in this work were conducted using the Cornell National Supercomputer Facility, a resource of the Center for Theory and Simulation in Science and Engineering at Cornell University, which is funded in part by the National Science Foundation, New York State, and the IBM Corporation, and members of the Corporate Research Institute.

J. W. Silverstein is with the Department of Mathematics, North Carolina State University, Raleigh, NC 27695.

P. L. Combettes is with the Department of Electrical Engineering, City College and Graduate School, City University of New York, New York, NY 10031.

IEEE Log Number 9201065.

of sensors. Common detection methods, such as those based on information theoretic criteria,[1] rely on the ergodic theorem. Their performance strongly depends on the spatial covariance matrix R of the data process being closely approximated by the sample covariance matrix $\hat{R}$, requiring the sample size to be quite large. In applications where the number of signals and, consequently, the number of sensors, is sizable, the required number of samples may be prohibitive. The purpose of this correspondence is to bring into play elements of the spectral theory of random matrices, more specifically, results on the limiting distribution of the eigenvalues of random matrices as the dimension increases. The analysis will show that, when the number of sensors is large, the number of signals can be estimated with a sample size considerably less than that required by invoking the ergodic theorem.

II. Spectral Theory of Random Matrices

Throughout this correspondence, all the random variables (r.v.'s) are defined on a probability space $(\Omega, \Sigma, \mathsf{P})$. The expressions i.d., i.i.d., and a.s. stand for identically distributed, independent and identically distributed, and P-almost surely, respectively. For distribution functions[2] (d.f.'s), $\Rightarrow$ denotes weak convergence. * will denote the set of positive integers and $^*_+$ the set of positive real numbers. The transpose of a matrix M is denoted by M^T and its conjugate transpose by M^*. Let M be an $m \times m$ random matrix with real-valued eigenvalues $(\Lambda_i)_{1 \le i \le m}$. The empirical d.f. of $(\Lambda_i)_{1 \le i \le m}$ is the stochastic process[3]

$$(\forall \omega \in \Omega)\ (\forall x \in \ )\ F^M(x, \omega) = \frac{1}{m} \sum_{i=1}^{m} 1_{]-\infty, x]}(\Lambda_i(\omega)). \qquad (1)$$

We now review the main result, a limit theorem found in [13].

Theorem 1: [13] Let $(Y_{ij})_{i,j \ge 1}$ be i.i.d. real-valued r.v.'s with $\mathsf{E}|Y_{11} - \mathsf{E}Y_{11}|^2 = 1$. For each m in *, let $Y_m = [Y_{ij}]_{m \times n}$, where $n = n(m)$ and $m/n \to y > 0$ as $m \to +\infty$, and let T_m be an $m \times m$ symmetric nonnegative definite random matrix independent of the Y_{ij}'s for which there exists a sequence of positive numbers $(\mu_k)_{k \ge 1}$ such that for each k in *

$$\int_0^{+\infty} x^k \, dF^{T_m}(x) = \frac{1}{m} \operatorname{tr} T_m^k \overset{\text{a.s.}}{\to} \mu_k \qquad \text{as } m \to +\infty \qquad (2)$$

and where the μ_k's satisfy Carleman's sufficiency condition, $\Sigma_{k \ge 1} \mu_{2k}^{-1/2k} = +\infty$, for the existence and the uniqueness of the d.f. H having moments $(\mu_k)_{k \ge 1}$. Let $M_m = (1/n) Y_m Y_m^T T_m$. Then, a.s., $(F^{M_m})_{m \ge 1}$ converges weakly to a nonrandom d.f. F having moments

$$(\forall k \in \ ^*)\ \nu_k = \sum_{w=1}^{k} y^{k-w} \sum \frac{k!}{m_1! \cdots m_w! \, w!} \mu_1^{m_1} \cdots \mu_w^{m_w} \qquad (3)$$

where the inner sum extends over all w-tuples of nonnegative integers $(m_1, \cdots, m_w)$ such that $\Sigma_{i=1}^{w} m_i = k - w + 1$ and $\Sigma_{i=1}^{w} i m_i = k$. Moreover, these moments uniquely determine F.[4] ◇

Although it does not appear likely a general explicit expression for F in terms of y and arbitrary H can be derived, useful qualitative information can be found from the different methods used in [6], [7], and [11] to express transforms of F (transforms of Stieltjes type in [6] and [11], the characteristic function in [7]). Of particular interest is the fact that the endpoints of the connected components of the support of F are given by the extrema of the function [6]

$$f(\alpha) = -\frac{1}{\alpha} + y \int_0^{+\infty} \frac{dH(x)}{\alpha + 1/x}. \qquad (4)$$

We now give new results apropos of the limiting behavior of $(F^{M_m})_{m \ge 1}$.

Theorem 2: [10] The limiting d.f. F in Theorem 1 is continuous on $^*_+$. Moreover, if H places no mass at 0 then, a.s., $(F^{M_m})_{m \ge 1}$ converges to F uniformly in . ◇

Proposition 1: [10] With the same notations and hypotheses as in Theorem 1, (i) F and y uniquely determine H; (ii) Almost surely, $(F^{T_m})_{m \ge 1}$ converges to H weakly; (iii) $F \Rightarrow H$ as $y \to 0$. ◇

Statement (iii) has a direct bearing on the problem of estimating the spectrum of a covariance matrix from observing that of a sample covariance matrix. Indeed, the matrix $(1/n) T_m^{1/2} Y_m Y_m^* T_m^{1/2}$ (whose eigenvalues are identical to those of M_m[5]) encompasses a broad class of sample covariance matrices stemming from n i.i.d. samples distributed as an m-dimensional random vector X with $\mathsf{E}X = 0$ and $\mathsf{E}XX^* = T_m$ (including the Wishart case when X is multivariate complex Gaussian). In estimating the spectrum of T_m from the sample covariance matrix, there seems to be no mention in the literature as to the dependence of n on m, that is, how large the sample size should be vis-à-vis the vector dimension in order to estimate the eigenvalues to within a certain degree of accuracy. Indeed, asymptotic results are expressed only in terms of the sample size (see, e.g., [1]). The fact that F differs from H for $y > 0$ while $F \Rightarrow H$ as $y \to 0$, which complements the fact that, for fixed m, $M_m \overset{a.s.}{\to} T_m$ as $n \to +\infty$, confirms the intuitively apparent statement that, for m large, n should be much larger, in the sense that $m = o(n)$.

III. Application to Signal Detection

A. Description of the Problem and Assumptions

Let p be the number of sensors in the array, q the unknown number of signals ($q < p$), and $[0, \tau]$ the observation interval. At each time t in $[0, \tau]$, the jth signal present in the scene, the additive noise at the ith sensor, and the received data at the ith sensor are, respectively, represented by the square-integrable complex-valued r.v.'s $S_j(t)$, $N_i(t)$, and $X_i(t)$. The random vectors $(S(t) = [S_1(t), \cdots, S_q(t)]^T)_{t \in [0, \tau]}$ are i.d. with nonsingular spatial covariance matrix $R_S = \mathsf{E}S(0)S(0)^*$. Moreover, it is assumed that the r.v.'s $(N_i(t) | 1 \le i \le p, t \in [0, \tau])$ are i.i.d. with $\mathsf{E}N_1(0) = 0$ and $\mathsf{E}|N_1(0)|^2 = \sigma^2$, where σ^2 is unknown, and independent from the r.v.'s $(S_j(t) | 1 \le j \le q, t \in [0, \tau])$. Let $N(t) = \sigma W(t) = \sigma[W_1(t) \cdots W_p(t)]^T$ (so that the $W_i(t)$'s are standardized) and $X(t) = [X_1(t) \cdots X_p(t)]^T$. The data collected by the array of sensors are modeled as observation of the random vector $X(t) = AS(t) + N(t)$, $t \in [0, \tau]$, where A is a $p \times q$ complex matrix depending on the geometry of the array and the parameters of the signals, and is assumed to have rank q.

The detection problem is to estimate q from the observation of n snapshots $(X(t_i))_{1 \le i \le n}$ of the data process. Under the above as-

[1] This approach was first proposed in [12] and further studied in [14], [16], and [17].

[2] By a d.f. we mean a right-continuous nondecreasing function F on with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$. The support of F is the closed set $\mathcal{S} = \{x \in \ | (\forall \epsilon \in \ ^*_+)\ F(x + \epsilon) > F(x - \epsilon)\}$.

[3] The indicator function of a set S is denoted by 1_S.

[4] Similar results are given in [6] and [11] with varying degrees of assumptions, although in both papers the matrices studied can have complex-valued entries. However, the proof in [13] can easily be modified to allow complex-valued entries in Y_m and T_m, giving the same result, provided T_m is Hermitian and we take $M_m = (1/n) Y_m Y_m^* T_m$.

[5] The reader is reminded that given two matrices $A_{p \times q}$ and $B_{q \times p}$, where $p \ge q$, the spectrum of AB is that of BA augmented by $p - q$ zeros.

sumptions, the random vectors $(X(t))_{t \in [0, \tau]}$ are i.d. with spatial covariance matrix $R = \mathsf{E}X(0)X(0)^* = AR_SA^* + \sigma^2 I_p$, where I_p denotes the $p \times p$ identity matrix. Moreover, the $p - q$ smallest eigenvalues of R are equal to σ^2. These eigenvalues will be referred to as the noise eigenvalues and the remainder of the spectrum will be referred to as the signal eigenvalues. In practice, R is not known, and its spectrum must be inferred from observing that of the sample covariance matrix $\hat{R} = (1/n) \Sigma_{i=1}^n X(t_i)X(t_i)^*$. Loosely speaking, one must then decide where the observed spectrum splits into noise and signal eigenvalues.

B. General Analysis

For every t in $[0, \tau]$, let us assume that the signal vector is given by

$$S(t) = CV(t) \quad \text{with} \quad V(t) = [V_1(t), \cdots, V_q(t)]^T \tag{5}$$

where C is $q \times q$, nonsingular, and the r.v.'s $(V_i(t))_{1 \le i \le q}$ are i.i.d. with the same d.f. as $W_1(0)$.[6] Let $B = AC$. Then

$$X(t) = [B \quad \sigma I_p] \begin{bmatrix} V(t) \\ W(t) \end{bmatrix}. \tag{6}$$

Notice that $R_S = CC^*$ and $R = BB^* + \sigma^2 I_p$. If we further assume that the n vectors $(S(t_i))_{1 \le i \le n}$ are independent, then the n data samples $(X(t_i))_{1 \le i \le n}$ will also be independent and the corresponding sample covariance matrix $\hat{R}$ takes on the form $\hat{R} = (1/n)[B \quad \sigma I_p] VV^*[B \quad \sigma I_p]^*$, where $V = [V_{ij}]_{(p+q) \times n}$ consists of i.i.d. standardized entries.

Theorem 3: [10] If $W_1(0)$ is standard complex Gaussian,[7] the joint distribution of the eigenvalues of $\hat{R}$ is the same as the joint distribution of the eigenvalues of $\hat{R}' = (1/n) Y_p Y_p^* (BB^* + \sigma^2 I_p)$, where Y_p is any $p \times n$ random matrix with i.i.d. standardized complex Gaussian entries. In general, for p and n sufficiently large, with high probability, the empirical d.f.'s $F^{\hat{R}}$ and $F^{\hat{R}'}$ are close to the d.f. F of Theorem 1 for $m = p$, $y = p/n$, and $H = F^{BB^* + \sigma^2 I_p}$. ◇

The importance of Theorem 3 becomes immediately apparent. The observations of the empirical d.f. $F^{\hat{R}}$, for suitably large p and n, will not vary very much from one realization to another, even if n is not large relative to p. In fact, by Theorem 2, with high probability, $F^{\hat{R}}$ will be uniformly close to a d.f. F that depends only on y and the eigenvalues of $BB^* + \sigma^2 I_p$. Hence, a realization of $F^{\hat{R}}$ and the ratio p/n can be used to describe, to within a certain degree of accuracy, $F^{BB^* + \sigma^2 I_p}$, which will yield σ^2 and the ratio $y_1 = q/p$ which corresponds to the q positive eigenvalues of BB^*.

Much of the information on the spectrum of $BB^* + \sigma^2 I_p$ can be directly observed from plotting histograms of the eigenvalues of $\hat{R}$, in particular, the ratio y_1 of signal eigenvalues. Let G denote the empirical d.f. of the eigenvalues of $BB^* + \sigma^2 I_p$ which are greater than σ^2, and let b_1 and b_2 denote, respectively, the smallest and largest of these values. Then, for every x in $\mathbb{R}$, we can write

$$H(x) = F^{BB^* + \sigma^2 I_p}(x) = (1 - y_1) 1_{[\sigma^2, +\infty[}(x) + y_1 G(x). \tag{7}$$

Proposition 2: [10] When $y < 1$, the smallest interval $[x_1, x_4]$ containing the support of F satisfies $0 < x_1 < x_4 < +\infty$ with $x_1 \uparrow \sigma^2$ and $x_4 \downarrow b_2$ as $y \downarrow 0$. In addition, there exists an α in $]-1/\sigma^2, -1/b_1[$ such that

$$g(\alpha) = y\left((1 - y_1)\left(\frac{\alpha}{\alpha + 1/\sigma^2}\right)^2 + y_1 \int_{b_1}^{b_2} \left(\frac{\alpha}{\alpha + 1/x}\right)^2 dG(x)\right) < 1 \tag{8}$$

(which can always be found for y sufficiently small) if and only if the support of F splits into at least two separate components, with the leftmost interval $[x_1, x_2]$ being a connected component of the support containing mass $1 - y_1$ from F. Furthermore, for y sufficiently small, $x_2 \downarrow \sigma^2$ as $y \downarrow 0$ and, if $[x_3, x_4]$ denotes the smallest interval containing the remaining support of F, then $x_3 \uparrow b_1$ as $y \downarrow 0$. Regardless of the respective location of x_2 and x_3 vis-à-vis σ^2 and b_1, the separation between the noise and signal portions of the spectrum, i.e., $x_3 - x_2$, increases as y decreases. When $y > 1$, F places mass $1 - 1/y$ at the origin, but the remaining support will lie to the right of a positive value x_1. It is still possible for the support of F to split further provided (8) holds. In this case the leftmost interval $[x_1, x_2]$ will carry mass $(1/y) - y_1$, leaving mass y_1 to the remaining support of F to the right of x_2. When $y = 1$ the latter situation applies, except now $x_1 = 0$, and there will be no mass at 0. ◇

Therefore, if p and n are large enough so that $F^{\hat{R}}$ is close to F with high probability, then for $y = p/n$ suitably small, an appropriately constructed histogram of the eigenvalues of $\hat{R}$ will display clustering on the left separated from the rest of the figure. The proportion of the number of eigenvalues associated with the histogram to the right of the clustering will then be close to q/p, with high probability.

Although the theory merely guarantees that the proportion of signal eigenvalues of $\hat{R}$ is close to that of R, extensive simulation strongly suggests that the spectrum of $\hat{R}$ splits into two portions containing the exact number of noise and signal eigenvalues, and that the endpoints of these portions agree very closely with the ones predicted by the theory. This point, which will be illustrated in Section IV, leads to the possibility of the existence of a much stronger underlying spectral theory deepening the results of Theorem 1. Results along these lines are known for the extreme eigenvalues when $T_m = \sigma^2 I_m$ and will be discussed Section III-C.

Intuitively, the above procedure has advantages over other methods used to estimate q, in particular, those adapted from information theoretic criteria [12], [14], [16], [17] which try to exploit the closeness of the noise eigenvalues of $\hat{R}$ to each other as well as their separation from the remaining signal eigenvalues. Usually the sample size has to be quite large for the smaller eigenvalues to cluster. On the other hand, only the separation of the two classes of eigenvalues is needed when viewing the spectrum, so a suitable n can conceivably be much smaller, sometimes even smaller than p. In other words, previous methods require $\hat{R}$ to be near $BB^* + \sigma^2 I_p$, while, for situations where p is sizable, the present analysis requires n to be large enough so that the support of F separates.

C. Specific Cases

An important case is the one for which no signal is present, that is, when $B = 0$, or equivalently, when $T_m = \sigma^2 I_m$. Then it is known [4]-[6] that, for $y \le 1$, F is continuously differentiable, where

$$F'(x) = \begin{cases} \dfrac{((x - \sigma^2(1 - \sqrt{y})^2)(\sigma^2(1 + \sqrt{y})^2 - x))^{1/2}}{2\pi\sigma^2 yx} \\ \qquad \text{if } \sigma^2(1 - \sqrt{y})^2 < x < \sigma^2(1 + \sqrt{y})^2 \\ 0 \quad \text{otherwise} \end{cases} \tag{9}$$

[6] It is worth noting that this general formulation comprises the special case when $S(0)$ is multivariate complex Gaussian, which is a common assumption in array signal processing.

[7] A r.v. is standardized complex Gaussian if its real and imaginary parts are i.i.d. with mean zero and variance $1/2$.

TABLE I
OBSERVED SPECTRA

	$y = 1$				$y = 1/5$				$y = 1/30$				$y = 0$
	$\mathfrak{L}_1$	$\mathfrak{L}_2$	$\mathfrak{L}_3$	$\mathfrak{L}_4$	$\mathfrak{L}_1$	$\mathfrak{L}_2$	$\mathfrak{L}_3$	$\mathfrak{L}_4$	$\mathfrak{L}_1$	$\mathfrak{L}_2$	$\mathfrak{L}_3$	$\mathfrak{L}_4$	$\mathfrak{L}$
λ_1	0.00	0.00	0.00	0.00	0.54	0.52	0.48	0.50	0.82	0.82	0.81	0.82	1
λ_2	0.00	0.01	0.01	0.01	0.60	0.56	0.55	0.57	0.84	0.84	0.84	0.86	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
λ_{10}	0.31	0.33	0.32	0.32	0.94	0.93	0.94	0.90	1.02	1.02	1.01	1.01	1
λ_{11}	0.43	0.40	0.37	0.41	1.01	0.98	1.00	0.97	1.04	1.04	1.03	1.03	1
λ_{12}	0.45	0.47	0.48	0.49	1.09	1.03	1.06	1.02	1.06	1.04	1.06	1.06	1
λ_{13}	0.57	0.57	0.50	0.60	1.12	1.06	1.13	1.11	1.08	1.07	1.09	1.08	1
λ_{14}	0.67	0.64	0.64	0.74	1.18	1.12	1.24	1.16	1.10	1.11	1.11	1.10	1
λ_{15}	0.86	0.87	0.83	1.05	1.36	1.27	1.31	1.32	1.16	1.13	1.13	1.14	1
λ_{16}	1.38	1.64	1.40	1.90	4.35	4.78	4.37	4.84	5.32	5.09	5.23	5.27	5.34
λ_{17}	2.59	2.72	2.41	2.81	5.26	6.05	6.31	6.13	5.97	6.34	6.08	6.10	6.20
λ_{18}	5.61	5.21	4.74	4.97	17.4	17.5	18.1	18.5	20.5	20.6	20.1	19.4	21.4
λ_{19}	7.98	7.64	8.22	7.37	19.4	20.1	18.7	20.4	22.0	22.1	21.7	21.3	23.1
λ_{20}	11.4	9.87	10.8	9.67	22.9	25.1	21.3	22.3	24.8	25.3	26.0	25.5	25.7
λ_{21}	14.8	13.3	11.9	11.7	36.7	40.2	39.6	39.8	48.6	48.8	49.6	47.7	49.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
λ_{49}	1159	1074	1137	1065	902	856	887	875	764	807	818	766	756
λ_{50}	1470	1309	1233	1390	1063	985	1004	1064	944	978	929	947	932

and for $y > 1$, F has derivative (9) on $\mathbb{R}_+^*$ and mass $1 - 1/y$ at 0. Furthermore, the largest eigenvalue of M_m converges a.s. (respectively, in probability) to $\sigma^2(1 + \sqrt{y})^2$ as $m \to +\infty$ if and only if $\mathsf{E}Y_{11} = 0$ and $\mathsf{E}|Y_{11}|^4 < +\infty$ (respectively, $x^4\mathsf{P}\{\omega \in \Omega \,|\, |Y_{11}(\omega)| \geq x\} \to 0$ as $x \to +\infty$] [2], [3], [9], [15]. The a.s. convergence of the smallest eigenvalue of M_m, to $\sigma^2(1 - \sqrt{y})^2$ when $y < 1$ has thus far been shown only for Y_{11} standardized Gaussian [8] (it is remarked here that the results on the extreme eigenvalues have been verified for Y_{11} real valued, but, again, the proofs can be extended to the complex case).

The above results can be used to investigate the possibility of no signals arriving at the sensors. Certainly, the existence of at least one signal would be in doubt if the number of samples was quite large but histograms indicate only one connected component away from 0. But, for any y, comparisons can be made between histograms of the eigenvalues, (9), and F' when $B \neq 0$, to infer whether or not signals are present, provided the latter densities exist and appear different enough from (9) to make a distinction. For this reason it is mentioned briefly here the case when G places mass at one value, $b > \sigma^2$. Except for the situation of only one signal, this case is not typically found in practice. However, the d.f. F can be completely determined and its properties strongly suggest the smoothness and appearance of F for general G. Only the case $y < 1$ will be outlined, the remaining cases for y following as above. From the analysis in [7] it can be shown that F is continuously differentiable with derivative of the form

$$F'(x) = \begin{cases} k\,\dfrac{(p_3(x) + x\sqrt{p_4(x)}^{1/3}) - (p_3(x) - x\sqrt{p_4(x)})^{1/3}}{x} \\ \qquad\qquad \text{if } p_4 \geq 0 \\ 0 \qquad \text{otherwise.} \end{cases} \tag{10}$$

Here, p_3 and p_4 are, respectively, third and fourth degree polynomials depending continuously on y, y_1, σ^2, b, and the leading coefficient of p_4 is negative. The latter polynomial has either two real roots, $0 < x_1 < x_4$, so that F has support on $[x_1, x_4]$, or four real roots, $0 < x_1 < x_2 < x_3 < x_4$, which is the above-mentioned case where the support of F splits into two intervals, $[x_1, x_2]$ and $[x_3, x_4]$, with $F(x_2) - F(x_1) = 1 - y_1$. Using (8), it is straightforward to show F splits if and only if

$$y\,\frac{((b^2y_1)^{1/3} + (\sigma^4(1 - y_1))^{1/3})^3}{(b - \sigma^2)^2} < 1. \tag{11}$$

When the left side of (11) is equal to 1, then p_4 still has four real roots, but $x_2 = x_3$. When (11) holds, F' is unimodal on each of the intervals, with infinite slopes at each endpoint. If there is a $y < 1$, say y_o, for which the left side of (11) is equal to 1, then, since the graph of F' varies continuously with y, as y increases from 0, the separate curves eventually join (at $y = y_o$) and the single curve will display two relative maxima, at least for y near y_o. Thus, although y may not be small enough to split F', it may still be possible to infer the number of signal eigenvalues from the shape of a histogram.

IV. Simulation Results

Our analysis applies to cases where p is large. Simulations have supported its applicability for values of p as low as 30. In the simulation presented here, a linear array with $p = 50$ sensors receives noisy signals from $q = 35$ narrow-band far-field sources. The sensors are assumed to be omnidirectional with unity gain and uniform spacing $\lambda/2$, where λ is the signal wavelength. The noise is zero mean, white, complex Gaussian, with power $\sigma^2 = 1$. The source signals are partially correlated with angles of arrivals uniformly spaced between $-70°$ and $70°$ and power selected at random from a uniform distribution so as to yield signal-to-noise ratios ranging from 0 to 10 dB. The signal vector is multivariate complex Gaussian and obtained according to (5), where C is a randomly generated banded matrix.

The spectrum $\mathfrak{L}$ of $R = BB^* + I_{50}$, where $B = AC$, was computed in order to obtain an explicit expression for the functions f of (4) and g of (8). Newton's method was used to find the minimum of $g(\cdot)$ over $]-1/\sigma^2, -1/b_1[$ and, whence, it was found that the largest value of y for which the splitting of the spectrum occurs

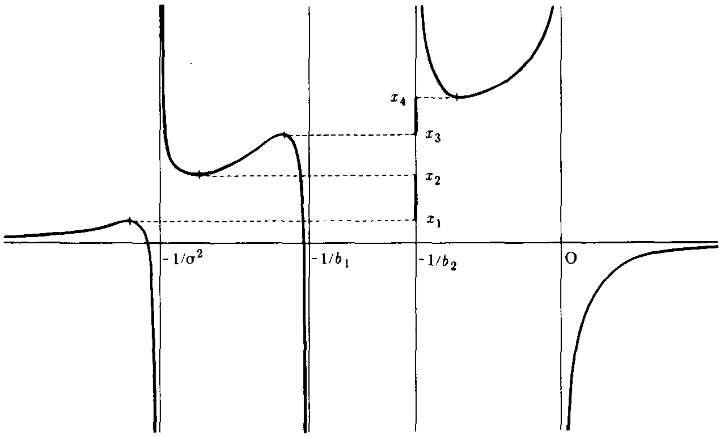

Fig. 1. Graph of f.

TABLE II
THEORETICAL BOUNDS FOR NOISE AND SIGNAL SPECTRUM SUPPORTS

	$y = 1$	$y = 1/5$	$y = 1/30$	$y = 0$
x_1	0.000	0.4642	0.789	1.000
x_2	1.124	1.369	1.184	1.000
x_3	1.167	3.970	5.785	5.342
x_4	1586	1137	995.4	931.6

(i.e., (8) holds) is $\bar{y} = 1.058$. Then, with the above configuration, three experiments were performed with the following number of samples n: 50, 250, and 1500 (which corresponds to values of y of 1, 1/5, and 1/30, respectively). In each experiment, several realizations $(\mathcal{L}_i)_i$ of the spectrum of the sample covariance matrix $\hat{R}$ were observed, the eigenvalues being arranged in nondecreasing order. Results indicating the variations of the observed spectra are shown in Table I. Even for $y = 1$, the $p - q = 15$ smallest eigenvalues are seen to cluster to the left of most of the observed spectra. This confinement delimitates exactly the noise portion of the spectrum and, thereby, detects the exact number of signals. As mentioned in Section II, for a given value of y, the theoretical end points x_1, x_2, and x_3, x_4, of the supports of the noise and signal portions of the spectrum can be determined from the location of the relative extrema of f in (4) (Fig. 1 shows a typical graph of f when separation occurs, as is the case here). Newton's method was used to this end and gave the results shown in Table II. In agreement with Proposition 2, it is seen that, as y decreases, the separation $x_3 - x_2$ increases while the endpoints converge towards the theoretical values.

V. CONCLUSIONS

This correspondence provides a theoretical analysis of the splitting of the spectrum of the sample covariance matrix between a connected noise component and a remaining signal component in situations where the number of sources is sizable. As far as the detection problem is concerned, the eigenvalues of the spatial covariance matrix R need not be estimated with a high degree of precision; only the accurate splitting of the spectrum is required. While conventional detection methods require that the sample size be impracticably large in order to closely approximate the spatial covariance matrix, the present analysis shows that the observed spectrum will split with high probability with a number of samples comparable to the number of sensors.

This work should suggest to the engineering community that by simply observing the spectrum of a large dimensional sample covariance matrix, highly relevant information can be extracted when the sample size is not exceedingly large. In the context of large dimensional array processing, its practical significance is that detection can be achieved when the sample size is only on the same order of magnitude as the number of sensors.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, second edition. New York: Wiley, 1984.
[2] Z. D. Bai, J. W. Silverstein, and Y. Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," *J. Multivariate Anal.*, vol. 26, no. 2, pp. 166-168, Aug. 1988.
[3] S. Geman, "A limit theorem for the norm of random matrices," *Ann. Probability*, vol. 8, no. 2, pp. 252-261, Apr. 1980.
[4] U. Grenander and J. W. Silverstein, "Spectral analysis of networks with random topologies," *SIAM J. Appl. Math.*, vol. 32, no. 2, pp. 499-519, Mar. 1977.
[5] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivariate Anal.*, vol. 12, no. 1, pp. 1-38, Mar. 1982.
[6] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR—Sbornik*, vol. 1, no. 4, pp. 457-483, 1967.
[7] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate *F* matrix," *SIAM J. Math. Anal.*, vol. 16, no. 3, pp. 641-646, May 1985.
[8] J. W. Silverstein, "The smallest eigenvalue of a large dimensional Wishart matrix," *Ann. Probability*, vol. 13, no. 4, pp. 1364-1368, Nov. 1985.
[9] J. W. Silverstein, "On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix," *J. Multivariate Anal.*, vol. 30, no. 2, pp. 307-311, Aug. 1989.
[10] J. W. Silverstein and P. L. Combettes, "Spectral theory of large dimensional random matrices applied to signal detection," Tech. Rep., Dep. Mathematics, North Carolina State Univ., Raleigh, NC, 1990.
[11] K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *Ann. Probability*, vol. 6, no. 1, pp. 1-18, Feb. 1978.
[12] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 387-392, Apr. 1985.
[13] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *J. Multivariate Anal.*, vol. 20, no. 1, pp. 50-68, Oct. 1986.
[14] Y. Q. Yin and P. R. Krishnaiah, "On some nonparametric methods

for detection of the number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 11, pp. 1533-1538, Nov. 1987.

[15] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probability Theory Related Fields*, vol. 78, no. 4, pp. 509-521, Aug. 1988.

[16] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, no. 1, pp. 1-25, Oct. 1986.

[17] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 2, pp. 129-132, Feb. 1987.

A Reverse Formulation of the RISE Algorithm

D. Mitchell Wilkes

Abstract—**The recently developed recursive/iterative self-adjoint eigenspace (RISE) decomposition algorithm recursively computes the complete eigenspace decomposition of successively larger Hermitian matrices. However, some practical applications also require that the decomposition of successively smaller matrices be computed. A modification to the RISE algorithm is presented that makes it possible to run this algorithm backward on successively smaller Hermitian matrices. This important modification increases the number of practical applications of this algorithm.**

I. Introduction

Recently, there has been an increasing amount of research in the area of recursive eigenspace decomposition. Beex and Fargues [1], [2] produced the numerically stable recursive/iterative Toeplitz eigenspace (RITE) decomposition. Wilkes and Hayes [3] and Morgera and Noor [4] then developed eigenvalue recursions for Toeplitz matrices, and although these algorithms were relatively efficient, they were also numerically ill behaved. Fargues and Beex extended RITE to colored noise problems (e.g., estimation of the direction of arrival of incident planes waves on equispaced linear arrays in the presence of colored noise) by developing the C-RITE algorithm [5]. C-RITE recursively computes the generalized eigenvalues and eigenvectors of successively larger Hermitian Toeplitz pencils. All of the techniques mentioned above are restricted to Hermitian Toeplitz matrices. This restriction was overcome by Wilkes via the recursive/iterative self-adjoint eigenspace (RISE) decomposition [6], which recursively finds the complete eigenspace decomposition of successively larger Hermitian matrices. Beex *et al.* then solved the generalized eigenspace decomposition for Hermitian pencils via the C-RISE algorithm [7].

RISE (as well as all of the other algorithms above) is limited to finding the eigenspace decomposition of successively larger matrices, with two possible ways in which the matrix may grow. The first way is the one described in [6] where a column and a row are added on the right and bottom of the matrix, respectively. That is, if the complete eigenspace decomposition of the $n \times n$ matrix R_n is known, RISE allows the efficient decomposition of R_{n+1} given by

$$R_{n+1} = \begin{bmatrix} R_n & r \\ r^* & r_0 \end{bmatrix} \tag{1}$$

where * indicates the conjugate transpose operation. Similarly, R_n may be allowed to grow by adding a row on the top and a column on the left.

$$R_{n+1} = \begin{bmatrix} r_0 & r^* \\ r & R_n \end{bmatrix} \tag{2}$$

This case is easily derived along the lines of the method given in [6]. Thus a sequence of successively larger R_n's may be allowed to evolve via any combination of these techniques, i.e., it is not necessary to restrict the growth to only one of the above two methods in this sequence. As a result, there are potentially many submatrices of any R_n for which the eigenspace decomposition is not known.

If it were possible to formulate a reverse version of RISE (so that the decomposition of successively smaller matrices may be found), it would be possible to find the decomposition of any of the Hermitian submatrices. Such a reverse formulation of RISE is presented in this letter. One application of this algorithm would be in the modeling of recursive input/output data (i.e., ARX modeling) using SVD or eigenspace-based techniques as in [8]. The observed input data, $x(n)$, and output data, $y(n)$, is assumed to obey a difference equation of the form

$$y(n) = -\sum_{k=1}^{p} a_k y(n-k) + \sum_{m=0}^{q} b_m x(n-m). \tag{3}$$

A combination of forward and reverse (or growing and shrinking) versions of RISE would enable the efficient estimation and comparison of the a_k and b_m parameters for many different values of p and q. Other applications include direction of arrival (DOA) estimation for large sensor arrays where the size and shape of a subarray is allowed to evolve, and beamspace formulations of the DOA problem where the number of beams used is allowed to grow (to improve detection performance) and shrink (to focus in on specific signals) [9].

For completeness, it is noteworthy to observe that the mathematics used in both the forward and reverse formulations of RISE are similar to those used in the rank-one update of eigenvalues [10] and eigenvalue perturbation analysis [11].

II. Reverse Formulation of Rise

It is assumed that the eigenspace decomposition of an $(n+1) \times (n+1)$ Hermitian matrix, R_{n+1}, is known. That is, R_{n+1} may be decomposed as

$$R_{n+1} = U_{n+1}\Lambda_{n+1}U_{n+1}^* \tag{4}$$

where the eigenvalues of Λ_{n+1} are ordered in the standard nonincreasing fashion, $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{n+1}$

$$\Lambda_{n+1} = \begin{bmatrix} \mu_1 & & \bigcirc \\ & \mu_2 & \\ & & \ddots \\ \bigcirc & & & \mu_{n+1} \end{bmatrix} \tag{5}$$

Manuscript received April 25, 1991; revised February 22, 1992.

The author is with the Department of Electrical Engineering, Vanderbilt University, Nashville, TN 37235.

IEEE Log Number 9201057.