

# LARGE DIMENSIONAL RANDOM MATRIX THEORY FOR SIGNAL DETECTION AND ESTIMATION IN ARRAY PROCESSING\*

*J. W. Silverstein*<sup>†</sup> and *P. L. Combettes*<sup>‡</sup>

<sup>†</sup>Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA.

<sup>‡</sup>Department of Electrical Engineering, City College and Graduate School,  
City University of New York, New York, NY 10031, USA.

## ABSTRACT

In this paper, we bring into play elements of the spectral theory of large dimensional random matrices and demonstrate their relevance to source detection and bearing estimation in problems with sizable arrays. These results are applied to the sample spatial covariance matrix,  $\widehat{R}$ , of the sensed data. It is seen that detection can be achieved with a sample size considerably less than that required by conventional approaches. As regards to determining the directions of arrivals, it is argued that more accurate estimates can be obtained by constraining  $\widehat{R}$  to be consistent with various *a priori* constraints, including those arising from large dimensional random matrix theory. A set theoretic formalism is used to formulate this feasibility problem. Unsolved issues are discussed.

## PROBLEM STATEMENT

We consider the problem of detecting the number  $q$  of sources impinging on an array of  $p$  ( $q < p$ ) sensors as well as their directions of arrival, when  $p$  is large. The model for the data formation mechanism is the following. At each time  $t$ , the  $j$ -th signal present in the scene, the additive noise at the  $i$ -th sensor, and the received data at the  $i$ -th 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  $(S(t) = [S_1(t) \dots S_q(t)]^T)_{t \in [0, +\infty[}$  are i.i.d. with  $ES(0) = 0$  and nonsingular spatial covariance matrix  $R_S = ES(0)S(0)^*$ . Moreover, it is assumed that the r.v.'s  $(N_i(t) \mid 1 \leq i \leq p, t \in [0, +\infty[)$  are i.i.d. with  $EN_1(0) = 0$  and  $E|N_1(0)|^2 = \sigma^2$ , where  $\sigma^2$  is unknown, and independent from the r.v.'s  $(S_j(t) \mid 1 \leq j \leq q, t \in [0, +\infty[)$ . Let  $N(t) = \sigma W(t) = \sigma[W_1(t) \dots W_p(t)]^T$  (so that the  $W_i(t)$ 's are standardized) and  $X(t) = [X_1(t) \dots X_p(t)]^T$ . The data collected by the array of

sensors are modeled as observations of the random vector  $X(t) = AS(t) + N(t)$ ,  $t \in [0, +\infty[$ , 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  $(X(t_i))_{1 \leq i \leq n}$  of the data process. Under the above assumptions, the random vectors  $(X(t))_{t \in [0, +\infty[}$  are i.i.d. with spatial covariance matrix  $R = EX(0)X(0)^* = AR_S A^* + \sigma^2 I_p$ , where  $I_p$  denotes the  $p \times p$  identity matrix. Moreover, the  $p - q$  smallest eigenvalues of  $R$  are equal to  $\sigma^2$ . These eigenvalues are referred to as the noise eigenvalues and the remainder of the spectrum is referred to as the signal eigenvalues. Since  $R$  is not known its spectrum must be inferred from observing that of the sample covariance matrix  $\widehat{R} = (1/n) \sum_{i=1}^n X(t_i)X(t_i)^*$ . Loosely speaking, one must then decide where the observed spectrum splits into noise and signal eigenvalues.

The estimation problem is to determine the direction of arrivals  $(\theta_i)_{1 \leq i \leq p}$  of the sources. Under standard hypotheses, this problem can be solved via the MUSIC method [11], which requires only the knowledge of  $R$ . In practice, however, only  $\widehat{R}$  is available, which leads to poor estimates if  $n$  is not sufficiently large.

In this paper, we bring into play elements of the spectral theory of large dimensional random matrices and demonstrate their relevance to source detection and estimation.

## SPECTRAL THEORY OF LARGE DIMENSIONAL RANDOM MATRICES

Let  $M$  be an  $m \times m$  random matrix with real-valued eigenvalues  $(\Lambda_i)_{1 \leq i \leq m}$ . Then, the empirical distribution function (d.f.) of  $(\Lambda_i)_{1 \leq i \leq m}$  is the stochastic process

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

\*The work of the first author was supported by NSF grant DMS-8903072.

We now review the main result, a limit theorem found in [15].

**Theorem 1.** [15] Let  $(Y_{ij})_{i,j \geq 1}$  be i.i.d. real-valued r.v.'s with  $\mathbb{E}|Y_{11} - \mathbb{E}Y_{11}|^2 = 1$ . For each  $m$  in  $\mathbb{N}^*$ , let  $Y_m = [Y_{ij}]_{m \times n}$ , where  $n = n(m)$  and  $m/n \rightarrow y > 0$  as  $m \rightarrow +\infty$ , 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  $(\mu_k)_{k \geq 1}$  such that for each  $k$  in  $\mathbb{N}^*$

$$\int_0^{+\infty} x^k dF^{T_m}(x) = \frac{1}{m} \text{tr} T_m^k \xrightarrow{\text{a.s.}} \mu_k \text{ as } m \rightarrow +\infty \quad (2)$$

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

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

where the inner sum extends over all  $w$ -tuples of nonnegative integers  $(m_1, \dots, m_w)$  such that  $\sum_{i=1}^w m_i = k - w + 1$  and  $\sum_{i=1}^w i m_i = k$ . Moreover, these moments uniquely determine  $F$ .<sup>1</sup>  $\diamond$

The following theorem pertains to the case when  $T_m$  is a multiple of the identity matrix.

**Theorem 2.** When  $T_m = \sigma^2 I_m$ ,  $F$  is known, having an algebraic density on the positive reals with support  $[\sigma^2(1 - \sqrt{y})^2, \sigma^2(1 + \sqrt{y})^2]$  [7], [8], [10]. The largest eigenvalue of  $M_m$  converges almost surely [respectively in probability] to  $\sigma^2(1 + \sqrt{y})^2$  as  $m \rightarrow +\infty$  if and only if  $\mathbb{E}Y_{11} = 0$  and  $\mathbb{E}|Y_{11}|^4 < +\infty$  [respectively  $x^4 \mathbb{P}\{|Y_{11}| \geq x\} \rightarrow 0$  as  $x \rightarrow +\infty$ ] [1], [6], [13], [16]. Moreover, if  $Y_{11}$  is standardized Gaussian, the smallest eigenvalue of  $M_m$  converges almost surely to  $\sigma^2(1 - \sqrt{y})^2$  when  $y < 1$  [12].<sup>2</sup>  $\diamond$

Several additional results on  $F$  are mentioned in [14], among them the convergence of  $F$  to  $H$  as  $y \rightarrow 0$ , and a way to compute the support of  $F$  from  $y$  and  $H$ .

## APPLICATION TO SIGNAL DETECTION

Existing approaches, such as those based on information theoretic criteria, rely on the closeness of the noise

<sup>1</sup>The proof in [15] 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  $M_m = (1/n)Y_m Y_m^* T_m$ .

<sup>2</sup>Results on the extreme eigenvalues have been verified for  $Y_{11}$  real-valued but, again, the proofs can be extended to the complex case.

eigenvalues of the sample spatial covariance matrix  $\widehat{R}$  to each other. This would require an extraordinarily large sample size (sometimes unattainable) when the number of sources is large in order to obtain a good estimate. Under additional assumptions on the signals (including the independence of the snapshots), Theorem 1 is used in [14] to show that, for  $p$  and  $n$  sufficiently large, with high probability, the empirical d.f.  $F^{\widehat{R}}$  is close to the d.f.  $F$  of Theorem 1 for  $m = p$ ,  $y = p/n$ , and  $H = F^{ARSA^* + \sigma^2 I_p}$ .

Further analysis when  $H$  is of this form shows that a value  $\tilde{y}$  can be computed for which  $y \in ]0, \tilde{y}[$  if and only if the support of  $F$  splits into at least two intervals, with the leftmost interval having mass  $(p - q)/p$ . For instance, in the simulations performed in [14],  $p = 50$  and  $\tilde{y}$  is found to be 1.058, which can allow a relatively small sample size. However, the simulations suggest something much stronger than the splitting of the support of  $F$  is occurring, namely exact splitting of the eigenvalues. Thus, upon mathematical verification of this phenomena the following appears valid:  $\widehat{R}$  with a sample size on the same order of magnitude as the number of sensors will have eigenvalues noticeably split into two groups, the group lying to the left corresponding to the multiplicity of the smallest eigenvalue of the true covariance matrix. Detection can thus be achieved with a sample size considerably less than that required by previous approaches.

## APPLICATION TO BEARING ESTIMATION

Under our basic assumptions, the directions of arrival can be calculated from the spatial covariance matrix  $R$  via the MUSIC algorithm [11]. In practice, short of knowing  $R$ , one must carry out the computation based on an observation of the sample covariance matrix  $\widehat{R}$ . Because  $\widehat{R}$  is often a poor approximation of  $R$ , the method can be improved by replacing  $\widehat{R}$  by a matrix that satisfies all the *a priori* constraints on the problem before applying MUSIC. By invoking the formalism of set theoretic estimation [3] and denoting by  $(\Psi_i)_{1 \leq i \leq M}$  the constraints, this feasibility problem can be formulated as that of finding a matrix  $\tilde{R}$  in the matrix subset

$$S = \bigcap_{i=1}^M S_i \text{ where } S_i = \{Q \mid Q \text{ satisfies } \Psi_i\}. \quad (4)$$

In general, finding a point in  $S$  directly is impossible. Let  $\Pi_i$  be the projection map onto  $S_i$ , i.e.,  $\Pi_i(Q)$  is the set of matrices in  $S_i$  which lie closest to  $Q$  in

terms of a distance (for computational tractability, we shall take the Frobenius distance). Then, under certain conditions on the sets and the initial point  $Q_0$ , the sequence  $(Q_n)_{n \geq 0}$  where  $Q_{n+1} \in \Pi_{i_n}(Q_n)$  with  $i_n = n \pmod{M} + 1$  will converge to a point  $\tilde{Q}$  in  $S$  [4]. In this scheme, the sets  $(S_i)_{1 \leq i \leq M}$  are activated in a cyclic manner and the update is any projection of the current estimate onto the next set.<sup>3</sup>

First of all we can pose the problem in the  $R$ -space and construct sets of estimates of the true covariance matrix. Under the above assumptions, an obvious *a priori* constraint about  $R$  is that its rank is  $q$ . One can therefore consider the (closed, nonconvex) set  $S_1$  of matrices whose rank is at most  $q$ . Other constraints may arise from the geometry of the array. Thus, if the array is linear with equispaced sensors,  $R$  will have a Toeplitz structure and one can take  $S_2$  to be the subspace of Toeplitz matrices. The use of a projection onto the set  $S_2$  has been reported in several array processing applications, e.g., [9] and is often referred to as Toeplitzation. The joint use of  $S_1$  and  $S_2$  via alternating projections was proposed in [2]. It should be noted that in such a process, loss of positive definiteness may occur. Therefore, a third set should be added, namely that of positive definite matrices. In simulations, noticeable improvements have been reported by using the constrained covariance matrix  $\tilde{R}$  instead of its raw counterpart  $\hat{R}$ , especially if the number of samples  $n$  and the signal-to-noise ratio are low.

In the above approach, one seeks to estimate  $R$  directly, which limits the exploitation of information that may be available about the noise. An alternative is to estimate the noiseless  $p \times n$  data matrix  $H = n^{-1/2}AS$  in the model  $X = AS + N$ .<sup>4</sup> An estimate  $\tilde{H}$  of  $H$  can be synthesized by projection onto various constraint sets. One can then form the constrained estimate of  $AR_S A^*$ , i.e.,  $\tilde{R} = \tilde{H}\tilde{H}^*$ , and apply MUSIC to it. Let us now consider the constraints that can be imposed on  $H$ . To this end, define the residual matrix for a given estimate  $\tilde{H}$  of  $H$  to be  $Y(\tilde{H}) = X - n^{1/2}\tilde{H}$ . Note that we have  $Y(H) = N$ . Therefore, all the statistical information pertaining to  $N$  can be imposed on  $\tilde{H}$  and several sets can be constructed according to this principle [5]. For instance, with our assumptions on the noise, the entries of  $Y(\tilde{H})$  should look like i.i.d. samples from a zero mean complex population with

<sup>3</sup>In particular, if all the sets are closed and convex in a Hilbert space, convergence to a feasible point is guaranteed for any  $Q_0$ .

<sup>4</sup>In this model, the  $i$ th ( $1 \leq i \leq n$ ) column of the matrices  $X$ ,  $S$ , and  $N$  are  $X(t_i)$ ,  $S(t_i)$  and  $N(t_i)$ , respectively.

second absolute moment  $\sigma^2$ . A direct application of the analysis of [5] to their sample mean leads to a set of the type

$$S_i = \{\tilde{H} \mid |\sum_{k=1}^{2np} y_k(\tilde{H})| \leq \delta_i\}, \quad (5)$$

where  $y(\tilde{H})$  is the vector obtained by stacking the real and imaginary parts of the entries of  $Y(\tilde{H})$ . Sets based on other statistics of the entries of  $Y(\tilde{H})$  can be obtained in a like-manner. This  $H$ -space framework also makes the use of constraints arising from the properties of large dimensional random matrices possible. Indeed, according to Theorem 2, a bound can be obtained on the largest (in the Gaussian case also on the smallest) singular value of  $Y(H)$ . In the case of the largest singular value, one obtains the set

$$S_i = \{\tilde{H} \mid \|Y(\tilde{H})\|_S^2 \leq n(\sigma^2(1 + \sqrt{y})^2 + \epsilon_i)\}, \quad (6)$$

where  $\|\cdot\|_S$  denotes the spectral norm and  $\epsilon_i$  reflects a confidence limit. Of course, all the constraints on  $AR_S A^*$  mentioned previously (rank, structure (e.g., Toeplitz), etc.) can still be used in the  $H$ -space. For example, the set associated with the Toeplitz constraint reads

$$S_i = \{\tilde{H} \mid \tilde{H}\tilde{H}^* \text{ is Toeplitz}\}. \quad (7)$$

## OPEN PROBLEMS

There are several mathematical questions that need to be addressed concerning the behavior of large dimensional random matrices in connection with applications to the above array processing problems. The three most relevant are outlined below.

**Extending Theorem 1.** The application of Theorem 1 requires two assumptions on the formation of the signal vector  $S(t)$ . The first is that  $S(t) = CV(t)$ , where  $C$  is a fixed  $q \times q$  nonsingular matrix, and  $V(t)$  is made up of i.i.d. random variables with same distribution as the noise components. Since signal and noise components are typically assumed to be Gaussian, this does not appear to be a major problem. The second assumption is the independence of the signal vector across snapshots. This is more serious, even though independent samples are assumed in several mathematical treatments (for example, the work on computing  $q$  from information theoretic criteria), and in most simulations found in the literature. The possibility of extending Theorem 1 to  $Y_m$  having stationary columns needs to be investigated.

**Eigenvalue Splitting.** In the detection problem, the observance of exact eigenvalue splitting in the simulations is striking. A limit property stronger than weak convergence of d.f.'s must be in effect, and a proof of this phenomenon should be pursued. The result is basically an extension of Theorem 2 on the extreme eigenvalues of  $(1/n)Y_m Y_m^T$ .

**Rate of Convergence.** On top of these problems is the general question of how fast the quantities are approaching their limiting values. The simulations performed in [14] show that for  $p = 50$  separation of the noise and signal eigenvalues of  $\hat{R}$  are in agreement with the support separation in  $F$ . Preliminary analysis on  $F^{(1/n)}Y_m Y_m^T$  suggests a rate of convergence of  $1/m$ , supporting the view that limiting behavior can be evidenced for  $m$  not exceedingly high.

Two additional problems are mentioned here.

**Computation of the Projections.** A shortcoming of the proposed set theoretic approach to the determination of the directions of arrivals is the numerical tedium sometimes involved in computing the projections at each iteration. In general, the sets are given in the form  $S_i = \{Q \mid g_i(Q) \leq \delta_i\}$ , where  $g_i$  is a given functional. A projection of a matrix  $Q'$  onto  $S_i$  is obtained by solving the minimization problem<sup>5</sup>

$$\min_Q \|Q' - Q\| \quad \text{subject to } g_i(Q) = \delta_i, \quad (8)$$

which can be approached via the method of Lagrange multipliers. However, in the case when  $S_i$  is not convex, local minima may occur. In such cases, efficient global minimization methods should be developed to solve (8).

**Convergence to a Feasible Point.** Due to the presence of nonconvex sets, the convergence of the successive projection algorithm to a feasible point cannot be guaranteed for any initial estimate [4]. Although starting the iterations at the point provided by the data (i.e.,  $R_0 = \hat{R}$  in the  $R$ -space approach or  $H_0 = n^{-1/2}X$  in the  $H$ -space approach) constitutes a sensible choice, it does not always ensure convergence. Therefore, the convergence issue deserves further investigation.

## REFERENCES

[1] Z. D. Bai, J. W. Silverstein, and Y. Q. Yin, "A Note on the Largest Eigenvalue of a Large Dimensional Sample Covariance Matrix," *Journal of Multivariate Analysis*, vol. 26, no. 2, pp. 166-168, August 1988.

[2] J. A. Cadzow, "Signal Enhancement - A Composite Property Mapping Algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49-62, January 1988.

[3] P. L. Combettes and M. R. Civanlar, "The Foundations of Set Theoretic Estimation," *ICASSP Proceedings*, pp. 2921-2924. Toronto, Canada, May 14-17, 1991.

[4] P. L. Combettes and H. J. Trussell, "Method of Successive Projections for Finding a Common Point of Sets in Metric Spaces," *Journal of Optimization Theory and Applications*, vol. 67, no. 3, pp. 487-507, December 1990.

[5] P. L. Combettes and H. J. Trussell, "The Use of Noise Properties in Set Theoretic Estimation," *IEEE Transactions on Signal Processing*, vol. 39, no. 7, pp. 1630-1641, July 1991.

[6] S. Geman, "A Limit Theorem for the Norm of Random Matrices," *The Annals of Probability*, vol. 8, no. 2, pp. 252-261, April 1980.

[7] U. Grenander and J. W. Silverstein, "Spectral Analysis of Networks with Random Topologies," *SIAM Journal on Applied Mathematics*, vol. 32, no. 2, pp. 499-519, March 1977.

[8] D. Jonsson, "Some Limit Theorems for the Eigenvalues of a Sample Covariance Matrix," *Journal of Multivariate Analysis*, vol. 12, no. 1, pp. 1-38, March 1982.

[9] J. P. Lecadre and P. Lopez, "Estimation d'une Matrice Interspectrale de Structure Imposée," *Traitement du Signal*, vol. 1, pp. 4-17, Décembre 1984.

[10] V. A. Marčenko and L. A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Mathematics of the USSR - Sbornik*, vol. 1, no. 4, pp. 457-483, 1967.

[11] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276-280, March 1986.

[12] J. W. Silverstein, "The Smallest Eigenvalue of a Large Dimensional Wishart Matrix," *The Annals of Probability*, vol. 13, no. 4, pp. 1364-1368, November 1985.

[13] J. W. Silverstein, "On the Weak Limit of the Largest Eigenvalue of a Large Dimensional Sample Covariance Matrix," *Journal of Multivariate Analysis*, vol. 30, no. 2, pp. 307-311, August 1989.

[14] J. W. Silverstein and P. L. Combettes, "Signal Detection via Spectral Theory of Large Dimensional Random Matrices," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, August 1992.

[15] Y. Q. Yin, "Limiting Spectral Distribution for a Class of Random Matrices," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 50-68, October 1986.

<sup>5</sup>As mentioned before, computational tractability seems to favor the use of the Frobenius norm.

- [16] 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 and Related Fields*, vol. 78, no. 4, pp. 509-521, August 1988.