On the consistency of likelihood penalization methods in large sensor networks

Pascal Vallet, Philippe Loubaton
Laboratoire d'informatique Gaspard Monge (IGM)
5 Bd. Descartes 77454 Marne-la-Vallée (France)
{vallet,loubaton}@univ-mlv.fr

Xavier Mestre
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Av. Carl Friedrich Gauss 08860 Castelldefels, Barcelona (Spain)
xavier.mestre@cttc.cat

***Abstract*—This paper is devoted to the problem of source detection with large sensor networks, in a context where the number of available samples N and the number of antennas M are of the same order of magnitude. We focus here on the popular likelihood penalization (LP) methods, such as Minimum Description Length (MDL) or Akaike Information Criterion (AIC). Such methods have been widely studied in the context where $N >> M$, and in particular the consistency of the MDL and the inconsistency of the AIC estimator were established in the asymptotic regime where $N \to \infty$ while M remains constant. We propose here an analysis in the asymptotic regime where M, N both converge to ∞ at the same rate, and using results from random matrix theory, we establish conditions on the penalty term to ensure consistency of LP methods in this latter regime. As a consequence, we deduce that the MDL method is always inconsistent while the AIC method can be consistent in certain situations.**

I. Introduction

The problem of detection, i.e, estimating the number of sources from noisy observations is a fundamental problem in signal processing, as well as in many other fields like geophysics, finance or biomedical engineering. In the context of array processing, detection is performed by using N observations collected by an array of M sensors. In general, detection is a first step to obtain more precise informations on source localization (Direction of Arrival estimation), power of noise and source signals, or to perform advanced techniques like source separation. Numerous methods have been proposed since the past 40 years, and in particular, likelihood penalization (LP) methods, which are among the most popular. These methods consist in estimating the number of source signals by computing the global maximum of a certain cost function, composed of a part due to the log-likelihood of the model and a penalty term. A lot of works have been devoted to build LP-based estimators, and especially the so-called Minimum Description Length (MDL) and Akaike Information Criteria (AIC) estimators (see Wax & Kailath [1]), which are now widely used in array processing. The statistical performance of these estimators have been mainly studied in the asymptotic regime where the number of available snapshots N converges to ∞ while the number of antennas M remains constant. In particular, it was shown by Zhao et al. [2] that the MDL estimator is consistent in the latter regime while the AIC estimator is inconsistent. In practical situations, LP methods are used for $N >> M$. However, it is not always possible to have such an amount of samples, e.g. due to stationarity constraints or when the number of antennas is large. In this context, it was recently proposed (see e.g. Mestre & Lagunas [3] in the context of DoA estimation) to consider the asymptotic regime in which M, N both converge to ∞ while the ratio $\frac{M}{N}$ converges to a positive constant. In this latter regime, LP methods, which rely on the eigenvalues of the sample covariance matrix of the observations, do not behave as usual, essentially because these sample eigenvalues exhibit a new behaviour, which is not taken into account.

The purpose of this paper is therefore to study how LP methods behave in this new asymptotic regime, and especially to provide a first order approximation of the likelihood term. This study will allow to deduce conditions on the penalty term which guaranty the consistency of LP methods. For this, we use recent random matrix theory results describing the behaviour of the largest sample eigenvalues, providing the fact that the number of sources to estimate is small compared to M.

We mention here that some studies have been devoted to study the performance of AIC and MDL estimator using tools from random matrix theory, see Nadler [4], Nadakuditi & Edelman [5], but none of these works addressed the consistency of LP methods in the asymptotic regime considered here.

The paper is organized as follows. In section II, we present the model of signals used in the remainder of the paper as well as the addressed problem. In section III, we summarize some results from random matrix theory, and especially results concerning the asymptotic behaviour of the largest sample eigenvalues. In section IV, we compute the asymptotic deterministic equivalent of LP cost function, and provide constraints of the penalty term to ensure consistency. Finally, we provide some numerical examples in section V.

II. Model and problem's statement

We consider K narrow-band source signals impinging on an array of M sensors, with $K < M$. At time n, the received signal $\mathbf{y}_n \in \mathbb{C}^M$ is represented by

$$\mathbf{y}_n = \mathbf{A}\mathbf{s}_n + \mathbf{v}_n, \tag{1}$$

This work was partially supported by the French agency Direction Générale de l'Armement (DGA) and by the Catalan Government (grant 2009SGR1046).

where $\mathbf{A}$ is the $M \times K$ matrix of steering-vectors, assumed full rank K, $\mathbf{s}_n$ is the signal transmitted by the K sources at time n and where $\mathbf{v}_n$ is an additive white Gaussian noise with covariance $\sigma^2\mathbf{I}$. In this paper, the signals $\mathbf{s}_n$ are i.i.d. accross n and follow a complex Gaussian distribution with zero mean and covariance matrix $\mathbf{\Gamma}$, so that $\mathbb{E}[\mathbf{y}_n\mathbf{y}_n^*] = \mathbf{A\Gamma A}^* + \sigma^2\mathbf{I}$. We stress the fact that the analysis provided in this paper could be easily extended to the case of unknown deterministic signals, by using recent results on large information-plus-noise random matrices (see e.g. the introduction of [6]). Assuming we collect $N > M$ samples of (1), the observed signal writes

$$\mathbf{Y}_N = \mathbf{A}\mathbf{S}_N + \mathbf{V}_N, \tag{2}$$

with $\mathbf{Y}_N = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$, $\mathbf{S}_N = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$ and $\mathbf{V}_N = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$. Under the additional assumption that $\mathbf{\Gamma}$ is full rank, the signal covariance matrix $\mathbf{A\Gamma A}^*$ has rank K. By denoting its eigenvalues $\lambda_{1,N}, \ldots, \lambda_{M,N}$ (in decreasing order), we have $\lambda_{1,N} \geq \ldots \geq \lambda_{K,N} > \lambda_{K+1,N} = \ldots = \lambda_{M,N} = 0$. In the same way, we denote by $\hat{\lambda}_{1,N} \geq \ldots \geq \hat{\lambda}_{M,N}$ the eigenvalues of the sample covariance matrix $\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}$.

The detection by LP methods consists in estimating the number of sources K by

$$\hat{K} = \underset{k=1,\ldots,M}{\operatorname{argmin}} \hat{J}_N(k), \tag{3}$$

where $\hat{J}_N(k)$ is a cost function given by

$$\hat{J}_N(k) = \hat{L}_N(k) + P_N(k), \tag{4}$$

In (4), $\hat{L}_N(k)$ is the log-likelihood (LL) term given by

$$\hat{L}_N(k) = \frac{1}{MN} \log \left[\frac{\left(\prod_{i=k+1}^{M} \hat{\lambda}_{i,N}\right)^N}{\left(\frac{1}{M-k}\sum_{i=k+1}^{M} \hat{\lambda}_{i,N}\right)^{N(M-k)}} \right], \tag{5}$$

and $P_N(k)$ a penalty term, which is given, for the AIC and MDL methods, respectively by [1] (see e.g. [1])

$$P_{\mathrm{AIC},N}(k) = -\frac{k(2M-k)}{MN}, \tag{6}$$
$$P_{\mathrm{MDL},N}(k) = -\frac{k(2M-k)\log(N)}{2MN}.$$

In the asymptotic regime where M is constant and $N \to \infty$, it was shown in [2] that, with probability one (w.p.1)

$$\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N} = \mathbf{A\Gamma A}^* + \sigma^2\mathbf{I} + \mathcal{O}\left(\sqrt{\frac{\log(\log(N))}{N}}\right),$$

which in turn allows to obtain the consistency of the estimate $\hat{K}$ defined in (3), i.e. $\hat{K} = K$ w.p.1 for N large. However, in the asymptotic regime where $M, N \to \infty$ such that $\frac{M}{N}$ converges to a positive constant,

$$\left\| \frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N} - \left(\mathbf{A\Gamma A}^* + \sigma^2\right) \right\|$$

does not converge to 0 anymore (Silverstein [7]), and the consistency of $\hat{K}$ is not guarantied.

[1] Notice that we normalized $\hat{L}_N(k)$ and $P_N(k)$ by $\frac{1}{MN}$, compared to [1]. Of course, it does not change the estimate $\hat{K}$, and it will be convenient for the asymptotic analysis provided in the next section.

III. Random matrix theory results

From now on, we assume the following asymptotic regime: $M = M(N)$ function of N so that $c_N = \frac{M}{N} \to c \in (0, 1)$ when $N \to \infty$. We first describe the behaviour of the sample eigenvalues in the case where there are no source signals (section III-A) and next study the case where a small number of source signals are received (section III-B).

A. Pure noise case

Under the assumption that no source signals are received by the array, we have $\mathbf{Y}_N = \mathbf{V}_N$. The global asymptotic behaviour of the sample eigenvalues is described through the empirical eigenvalue distribution of $\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}$, defined as the random distribution

$$\hat{\mu}_N = \frac{1}{M}\sum_{k=1}^{M} \delta_{\hat{\lambda}_{k,N}}.$$

where δ_x is the Dirac measure at x. The behaviour of $\hat{\mu}_N$ is given by the following result.

Theorem 1 (Marcenko & Pastur [8])**.** *There exists a deterministic distribution μ, called the Marcenko-Pastur distribution, such that*

$$\hat{\mu}_N \xrightarrow[N\to\infty]{\mathcal{D}} \mu, \quad \textit{w.p.1}, \tag{7}$$

having density given by

$$\frac{\mathrm{d}\mu(x)}{\mathrm{d}x} = \frac{\sqrt{(x-x^-)(x^+-x)}}{2\pi\sigma^2 cx}\mathbb{1}_{[x^-,x^+]}(x),$$

where $x^- = \sigma^2(1-\sqrt{c})^2$ and $x^+ = \sigma^2(1+\sqrt{c})^2$.

In addition to this result, we have (Bai & Silverstein [9]), for each $\epsilon > 0$, that $\hat{\lambda}_{1,N}, \ldots, \hat{\lambda}_{M,N} \in [x^- - \epsilon, x^+ + \epsilon]$ w.p.1 for N large, that is, asymptotically, no sample eigenvalue escapes outside the compact support $[x^-, x^+]$ of the Marcenko-Pastur distribution. This result implies in particular that $\hat{\lambda}_{1,N} \to x^+$ and $\hat{\lambda}_{M,N} \to x^-$ w.p.1 as $N \to \infty$.

B. Perturbation by a small number of sources

We now evaluate the effect of the presence of a small number of signals, and we consider $\mathbf{Y}_N = \mathbf{A}\mathbf{S}_N + \mathbf{V}_N$, with $K = \operatorname{rank}(\mathbf{A\Gamma A}^*)$ independent of N. In this case, $\mathbf{Y}_N$ is of course a multiplicative small rank perturbation of the pure noise model, since

$$\mathbf{Y}_N \overset{\mathcal{D}}{=} \left(\mathbf{A\Gamma A}^* + \sigma^2\mathbf{I}\right)^{1/2}\mathbf{X}_N,$$

with $\mathbf{X}_N$ having independent standard Gaussian entries.

By standard techniques, it is possible to show that we still have $\hat{\mu}_N \xrightarrow{\mathcal{D}} \mu$ w.p.1. as $N \to \infty$. Hence, a fixed number of sources does not modify the global asymptotic behaviour of the sample eigenvalues. The difference with the pure noise case will naturally occur in the behaviour of the largest eigenvalues, which is given by the following result.

Theorem 2 (Baik et al. [10]). *Assume that $\lambda_{1,N}, \dots, \lambda_{K,N}$ converge to $\lambda_1, \dots, \lambda_K$. Then, for $k = 1, \dots, K$*

$$\hat{\lambda}_{k,N} \xrightarrow[N\to\infty]{a.s.} \psi(\lambda_k) \triangleq \begin{cases} x^+ & \text{if } \lambda_k \leq \sigma^2\sqrt{c}, \\ \frac{(\lambda_k+\sigma^2)(\lambda_k+\sigma^2 c)}{\lambda_k} & \text{if } \lambda_k > \sigma^2\sqrt{c}. \end{cases} \tag{8}$$

and $\hat{\lambda}_{K+1,N} \xrightarrow[N\to\infty]{a.s.} x^+$.

Thus, if λ_K is above the threshold $\sigma^2\sqrt{c}$, exactly K sample eigenvalues will escape from the support $[x^-, x^+]$ of the Marcenko-Pastur distribution. Consequently, the condition $\frac{\lambda_K}{\sigma^2} > \sqrt{c}$ can be seen as a detectability threshold, in the sense that is ensures asymptotic separation between "signal" and "noise" sample eigenvalues. It can be also interpreted as a SNR threshold, if we define the SNR as $\frac{\lambda_K}{\sigma^2}$. The escape of the K largest sample eigenvalues is illustrated in figure 1.

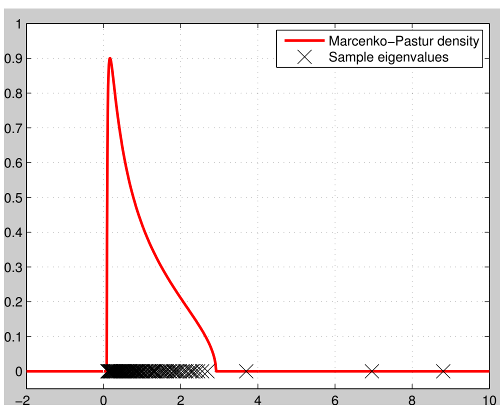

Figure 1. Density of the Marcenko-Pastur distribution and samples eigenvalues (black crosses), with $M = 100$, $N = 200$, $\sigma = 1$, $K = 3$ and signal eigenvalues $\lambda_{1,N} = 2$, $\lambda_{2,N} = 5$, $\lambda_{3,N} = 7$.

IV. Asymptotic behaviour of LP methods

In this section, we use the previous results on the behaviour of the largest sample eigenvalues to study first order approximation of the log-likelihood $\hat{L}_N(k)$ given in (5), and then deduce conditions on the penalty term $P_N(k)$ to ensure consistency. We therefore assume that $M = M(N)$ is a function of N so that $c_N = \frac{M}{N} \to c \in (0,1)$ when $N \to \infty$, and K is independent of N. From now on, we define the function $f(x) = x - \log(x)$.

Proposition 1. *Assume the conditions of Theorem 2 and set $\lambda_{K+1} = 0$. Then we have, for $k = 1, \dots, K+1$, w.p.1,*

$$\begin{aligned}\hat{L}_N(k) =& \\ &\frac{c-1}{c}\log(1-c) - 1 \\ &+ \frac{1}{M}\sum_{i=1}^{k}\left[f\left(\frac{\psi(\lambda_i)}{\sigma^2}\right) - 1\right] + \delta_N + o\left(\frac{1}{N}\right),\end{aligned}$$

where $\delta_N = o(1)$ a.s. and does not depend on k.

Proof: We first express the LL, for $k = 1, \dots, K+1$, as

$$\hat{L}_N(k) = \frac{1}{M}\sum_{i=k+1}^{M}\log\left(\hat{\lambda}_{i,N}\right) - \frac{M-k}{M}\log\left(\frac{1}{M-k}\sum_{i=k+1}^{M}\hat{\lambda}_{i,N}\right). \tag{9}$$

Using Theorem 2 and a Taylor expansion of the logarithm, the first term on the right-hand side (r.h.s) of (9) writes w.p.1

$$\frac{1}{M}\sum_{i=k+1}^{M}\log\left(\hat{\lambda}_{i,N}\right) = \frac{1}{M}\log\det\left(\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}\right) - \frac{1}{M}\sum_{i=1}^{k}\log\left(\psi(\lambda_i)\right) + o\left(\frac{1}{N}\right).$$

We clearly have

$$\frac{1}{M}\log\det\left(\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}\right) = \int_{\mathbb{R}}\log(x)\mathrm{d}\hat{\mu}_N(x), \tag{10}$$

and from the results of Section III-A and III-B, (10) converges to $\int_{x^-}^{x^+}\log(x)\mathrm{d}\mu(x)$ w.p.1. as $N \to \infty$. From Bai & Silverstein [11],

$$\int_{x^-}^{x^+}\log(x)\mathrm{d}\mu(x) = \log(\sigma^2) + \frac{c-1}{c}\log(1-c) - 1,$$

and thus

$$\begin{aligned}\frac{1}{M}\sum_{i=k+1}^{M}\log\left(\hat{\lambda}_{i,N}\right) =& \\ &\log(\sigma^2) + \frac{c-1}{c}\log(1-c) - 1 \\ &- \frac{1}{M}\sum_{i=1}^{k}\log\left(\psi(\lambda_i)\right) + \epsilon_N + o\left(\frac{1}{N}\right),\end{aligned} \tag{11}$$

where $\epsilon_N = o(1)$ w.p.1 and is independent of k. Similarly, the second term on the r.h.s. of (9) writes

$$\begin{aligned}\log\left(\frac{1}{M-k}\sum_{i=k+1}^{M}\hat{\lambda}_{i,N}\right) =& \\ &\log\left(\frac{1}{M}\mathrm{Tr}\left(\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}\right)\right) \\ &+ \frac{k}{M-k} - \frac{\frac{1}{M-k}\sum_{i=1}^{k}\psi(\lambda_i)}{\frac{1}{M}\mathrm{Tr}\left(\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}\right)} + o\left(\frac{1}{N}\right).\end{aligned}$$

As previously, we also have

$$\frac{1}{M}\mathrm{Tr}\left(\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N}\right) \xrightarrow[N\to\infty]{a.s.} \int_{x^-}^{x^+} x\mathrm{d}\mu(x).$$

But it also holds that

$$\frac{1}{M}\mathrm{Tr}\left(\frac{\mathbf{V}_N\mathbf{V}_N^*}{N}\right) \xrightarrow[N\to\infty]{a.s.} \int_{x^-}^{x^+} x\mathrm{d}\mu(x),$$

and the law of large number implies $\int_{x^-}^{x^+} x \mathrm{d}\mu(x) = \sigma^2$. Thus,

$$\log\left(\frac{1}{M-k}\sum_{i=k+1}^{M}\hat{\lambda}_{i,N}\right) = \log\left(\sigma^2\right) + \frac{k}{M-k} - \frac{1}{M-k}\sum_{i=1}^{k}\frac{\psi(\lambda_i)}{\sigma^2} + \epsilon'_N + o\left(\frac{1}{N}\right). \tag{12}$$

with $\epsilon'_N = o(1)$ w.p.1 independent of k. Gathering the estimates (11) and (12) with (9), we obtain after simple algebra the desired result. ∎

As a first immediate consequence of Proposition 1, we see that the MDL estimator is not consistent. Indeed, in the LL term $\hat{L}_N(k)$, the information about k is carried by a $\mathcal{O}\left(\frac{1}{N}\right)$ term and is overshadowed by the MDL penalty term which is $\mathcal{O}\left(\frac{\log(N)}{N}\right)$, for N large.

Since the AIC penalty term is $\mathcal{O}\left(\frac{1}{N}\right)$, we need precise constraints on the penalty term $P_N(k)$, to guaranty consistency of LP methods, which are given in the following result, whose proof is straightforward and therefore omitted.

Corollary 1. *Assume the conditions of Theorem 2 and that all the sources are asymptotically detectable, i.e.* $\lambda_K > \sigma^2\sqrt{c}$. *Let* $0 < \eta < f\left(\frac{\psi(\lambda_K)}{\sigma^2}\right) - f\left((1+\sqrt{c})^2\right)$ *and*

$$P_N(k) = -\frac{k}{M}\Big(f\left((1+\sqrt{c})^2\right) - 1 + \eta\Big). \tag{13}$$

Then w.p.1, for all large N,

$$\hat{J}_N(0) < \ldots < \hat{J}_N(K) \text{ and } \hat{J}_N(K) > \hat{J}_N(K+1).$$

By defining

$$\hat{K} = \min\left\{k : \hat{J}_N(k) > \hat{J}_N(k+1)\right\}, \tag{14}$$

Corollary 1 thus implies that w.p.1, for N large, $\hat{K} = K$. The fact that all the sources are detectable, i.e. $\lambda_K > \sigma^2\sqrt{c}$, is of course necessary (but not sufficient) for the consistency of LP methods, and it can be easily checked that if this condition is not verified, then $\hat{J}_N(k)$ will have a local maximum at the point $\max\left\{k : \lambda_k > \sigma^2\sqrt{c}\right\}$ (or at 0 if no sources are detectable), and therefore K will be underestimated w.p.1. for N large.

Comparing the penalty term of the AIC (6), which is equal to $-\frac{2kc}{M} + o\left(\frac{1}{N}\right)$, to (13), we deduce that the AIC estimator defined by (14) with penalty term (6), is consistent iff

$$1 + 2c < f\left(\frac{\psi(\lambda_K)}{\sigma^2}\right),$$

a condition which holds iff $\frac{\psi(\lambda_K)}{\sigma^2}$ is large enough.

V. Numerical Examples

In this section, we illustrate the performance of the MDL and AIC estimator in terms of probability of misdetection. We consider a ratio $c_N = c = 0.5$, a noise variance $\sigma^2 = 1$ and a matrix $\mathbf{A}\Gamma\mathbf{A}^*$ having $K = 3$ non-zero eigenvalues at $\lambda_{1,N} = \sigma^2\sqrt{c}+4$, $\lambda_{2,N} = \sigma^2\sqrt{c}+3$ and $\lambda_{3,N} = \sigma^2\sqrt{c}+0.3$.

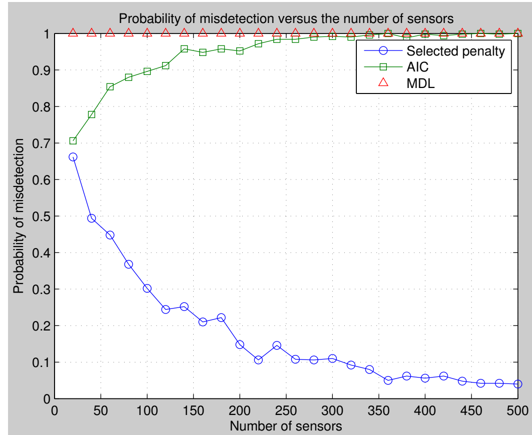

Figure 2. Emp. prob. of misdetection (1000 trials) vs number of sensors

In this case, $f\left(\frac{\psi(\lambda_3)}{\sigma^2}\right) < 1 + 2c$ and consistency of the AIC is not guarantied. Figure 2 shows the empirical probability of misdetection vs the number of sensors (with $N = \frac{M}{c} = 2M$ samples), for the AIC and MDL estimator, as well as a generic LP estimator having a penalty term as in (13) selected with

$$\eta = 0.05\left(f\left(\frac{\psi(\lambda_3)}{\sigma^2}\right) - f\left((1+\sqrt{c})^2\right)\right).$$

Both MDL and AIC estimator are not consistent, while the LP estimator with selected penalty term is consistent, thus validating Corollary 1. We stress the fact that the selected LP estimator is used to illustrate Corollary 1, and is of course not usable in practice, since η is not available.

References

[1] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.

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

[3] X.Mestre and M.A.Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, p. 598, 2008.

[4] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, 2009.

[5] R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, 2008.

[6] P. Loubaton and P. Vallet, "Almost sure localization of the eigenvalues in a gaussian information plus noise model. applications to the spiked models," *Electron. J. Probab.*, vol. 16, pp. 1934–1959.

[7] J. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, 1995.

[8] V. Marcenko and L. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sb.*, vol. 1, pp. 457–483, 1967.

[9] Z. Bai and J. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.

[10] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.

[11] Z. Bai and J. Silverstein, "Clt for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, no. 1A, pp. 553–605, 2004.