

On the behaviour of the estimated fourth-order cumulants matrix of a high-dimensional gaussian white noise

Pierre Gouédard, Philippe Loubaton

► **To cite this version:**

Pierre Gouédard, Philippe Loubaton. On the behaviour of the estimated fourth-order cumulants matrix of a high-dimensional gaussian white noise. LVA-ICA , Feb 2017, Grenoble, France. <hal-01616290>

HAL Id: hal-01616290

<https://hal-upec-upem.archives-ouvertes.fr/hal-01616290>

Submitted on 13 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the behaviour of the estimated fourth-order cumulants matrix of a high-dimensional Gaussian white noise

Pierre Gouédard and Philippe Loubaton *

Université Paris-Est Marne la Vallée, Laboratoire d'Informatique Gaspard Monge,
UMR CNRS 8049

5 Bd. Descartes, Cité Descartes, 77454 Marne la Vallée Cedex 2, France.

pierre.mgouedard@gmail.com

loubaton@univ-mlv.fr

Abstract. This paper is devoted to the study of the traditional estimator of the fourth-order cumulants matrix of a high-dimensional multivariate Gaussian white noise. If M represents the dimension of the noise and N the number of available observations, it is first established that this $M^2 \times M^2$ matrix converges towards 0 in the spectral norm sense provided $\frac{M^2 \log N}{N} \rightarrow 0$. The behaviour of the estimated fourth-order cumulants matrix is then evaluated in the asymptotic regime where M and N converge towards $+\infty$ in such a way that $\frac{M^2}{N}$ converges towards a constant. In this context, it is proved that the matrix does not converge towards 0 in the spectral norm sense, and that its empirical eigenvalue distribution converges towards a shifted Marcenko-Pastur distribution. It is finally claimed that the largest and the smallest eigenvalue of the cumulant matrix converges almost surely towards the rightend and the leftend points of the support of the Marcenko-Pastur distribution.

Keywords: Estimated joint fourth-order cumulants matrices, large random matrices, blind source separation in the high-dimensional context.

1 Introduction

It is now well understood that the statistical signal processing of high-dimensional signals poses a number of new problems which stimulated the development of appropriate new tools, e.g. large random matrices or approaches exploiting sparsity. Statistical methods based on the use of the empirical covariance matrix of a M -dimensional time series $(\mathbf{y}_n)_{n \in \mathbb{Z}}$ (e.g. detection of noisy low rank signals, estimation of direction of arrival using subspace methods,...) provide a number of convincing examples illustrating this point. In effect, when the dimension M of the observation is large, it is very often difficult to collect a number N of observations much larger than M , so that in practice M and N appear to be of the same order of magnitude. In this context, it is well established that

* This work was supported by project ANR-12-MONU-0003 DIONISOS.

the empirical covariance matrix $\hat{\mathbf{R}}_N = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*$ is a poor estimator of $\mathbf{R} = \mathbb{E}(\mathbf{y}_n \mathbf{y}_n^*)$. When no a priori information on \mathbf{R} (e.g. various kinds of sparsity, see e.g. [3]) is available, it appears that large random matrix theory provides useful informations on the behaviour of $\hat{\mathbf{R}}_N$ in the asymptotic regime where M and N converge towards $+\infty$ at the same rate. This allows to analyse the behaviour of the standard statistical inference algorithms based on the assumption that $\hat{\mathbf{R}}_N \simeq \mathbf{R}$, and more importantly to propose modifications which allow to improve the performance (see e.g. [2] [6], [7], [8], [14]).

The present paper is motivated by the blind source separation of an instantaneous mixture of K independent sources in the case where the observed signal $(\mathbf{y}_n)_{n \in \mathbb{Z}}$ is high-dimensional and where the additive noise is Gaussian with unknown statistics. Popular approaches developed and evaluated in the low dimensional observation context use the particular structure of the fourth-order cumulants tensor of the observed signal which appears as the sum of K rank 1 tensors generated by the columns of the mixing matrix. Under mild assumptions, the column vectors of the mixing matrix can be identified from the eigenvalue / eigenvector decomposition of the $M^2 \times M^2$ fourth-order cumulants matrix (see e.g. the algorithm ICAR in [1]). In practice, the fourth-order cumulants matrix has to be estimated from the N available M -dimensional observations $\mathbf{y}_1, \dots, \mathbf{y}_N$, and the presence of the additive Gaussian noise has of course an influence on the eigenstructure of the estimated fourth-order cumulants matrix and thus on the statistical performance of the estimator of the mixing matrix. When the dimension of the observation M is much smaller than the sample size N , standard large sample analysis conducted in the regime M fixed and $N \rightarrow +\infty$ can be used in order to prove that the estimated fourth-order cumulants matrix converges towards the true cumulants matrix in the spectral norm sense, a property that immediately implies the consistency of the mixing matrix estimates. When M is large, the above regime may not be relevant, and asymptotic regimes for which both M and N converge towards $+\infty$ at possibly different rates may produce more reliable results. In this context, a crucial issue is to determine the rates of convergence of M and N towards $+\infty$ for which the estimated fourth-order cumulants matrix still converges towards the true cumulant matrix in the spectral norm sense. When these conditions are not met, the traditional estimates of the mixing matrix are non consistent, but it may be useful to characterize the properties of the estimated fourth-order cumulants matrix in order to be able to derive improved performance estimates. This research program appears highly non trivial, and needs to develop a number of new large random matrix tools. In this paper, we thus consider the preliminary problem of characterizing the behaviour of the estimated fourth-order cumulants matrix in the absence of source when both M and N converge towards $+\infty$. We do not claim that the results of this paper can be used as is in order to analyse the behaviour of blind source separation algorithms in the high-dimensional case. However, the study of this simpler problem will provide a number of useful insights to address the more complicated scenario in which sources are present.

This paper is organized as follows. In section 2, we present more precisely the addressed problem. In section 3, we study the conditions on M and N under which the estimated fourth-order cumulants matrix of the M -dimensional white Gaussian noise sequence $(\mathbf{y}_n)_{n \in \mathbb{Z}}$ converges towards 0 in the spectral norm sense. We show that this is the case as soon as $\frac{M^2 \log N}{N} \rightarrow 0$, a condition which is close from $\frac{M^2}{N} \rightarrow 0$. In section 4, we consider the regime in which $\frac{M^2}{N}$ converges towards a non zero constant, and prove that the estimated fourth-order cumulants matrix does not converge towards 0 in the spectral norm sense. In particular, we establish that its empirical eigenvalue distribution converges towards a shifted Marcenko-Pastur (MP) distribution, and that its smallest and largest eigenvalues converge towards the leftend and rightend points of the support of the MP distribution. This result shows that parameter $\frac{M^2}{N}$ controls the spreading of the eigenvalues, thus confirming the evaluations of section 3. Some numerical experiments illustrating these results are also provided.

General notations. $(\mathbf{e}_i)_{i=1,\dots,M}$ represents the canonical basis of \mathbb{C}^M . If \mathbf{x} is an element of \mathbb{C}^{M^2} , and if $(k, i) \in \{1, 2, \dots, M\}$, we denote its $k + i(M - 1)$ -th component as $\mathbf{x}_{i,k}$. If \mathbf{A} and \mathbf{B} are 2 matrices, $\mathbf{A} \otimes \mathbf{B}$ represents the block matrix whose blocks are the $\mathbf{A}_{i,j} \mathbf{B}$. If \mathbf{A} is a $M^2 \times M^2$ matrix, we denote by $\mathbf{A}_{i,k,j,l}$ the element $k + i(M - 1), l + M(j - 1)$ of matrix \mathbf{A} . In other words, matrix \mathbf{A} can be written as

$$\mathbf{A} = \sum_{i,j,k,l=1,\dots,M} \mathbf{A}_{i,k,j,l} (\mathbf{e}_i \otimes \mathbf{e}_k) (\mathbf{e}_j \otimes \mathbf{e}_l)^*$$

In the following, we denote by $\mathbf{\Pi}$ the $M^2 \times M^2$ matrix defined by $(\mathbf{\Pi} \mathbf{x})_{i,j} = \mathbf{x}_{j,i}$ for each element \mathbf{x} of \mathbb{C}^{M^2} and for each pair (i, j) . It is clear that for each pair $(\mathbf{x}_1, \mathbf{x}_2)$ of \mathbb{C}^M , then it holds that $\mathbf{\Pi}(\mathbf{x}_1 \otimes \mathbf{x}_2) = \mathbf{x}_2 \otimes \mathbf{x}_1$. It is easily seen that matrix $\frac{1}{2}(\mathbf{I} + \mathbf{\Pi})$ can be expressed as

$$\frac{1}{2}(\mathbf{I} + \mathbf{\Pi}) = \mathbf{\Gamma} \mathbf{\Gamma}^*$$

where $\mathbf{\Gamma}$ is the $M^2 \times M(M + 1)/2$ matrix whose columns $\mathbf{\Gamma}_{i,j}$, $1 \leq i \leq j \leq M$ are $\mathbf{\Gamma}_{i,i} = \mathbf{e}_i \otimes \mathbf{e}_i$ and, for $i < j$,

$$\mathbf{\Gamma}_{i,j} = \frac{1}{\sqrt{2}} (\mathbf{e}_i \otimes \mathbf{e}_j + \mathbf{e}_i \otimes \mathbf{e}_j)$$

Matrix $\mathbf{\Gamma}$ verifies $\mathbf{\Gamma}^* \mathbf{\Gamma} = \mathbf{I}_{M(M+1)/2}$. Therefore, for each pair $(\mathbf{x}_1, \mathbf{x}_2)$ of \mathbb{C}^M , it holds that

$$\mathbf{\Gamma} \mathbf{\Gamma}^* (\mathbf{x}_1 \otimes \mathbf{x}_2) = \frac{1}{2} (\mathbf{x}_1 \otimes \mathbf{x}_2 + \mathbf{x}_2 \otimes \mathbf{x}_1) \quad (1)$$

2 Presentation of the problem.

We consider a sequence $(\mathbf{y}_n)_{n=1,\dots,N}$ of complex Gaussian i.i.d. $\mathcal{N}_c(0, \sigma^2 \mathbf{I}_M)$ random vectors. The fourth order joint cumulants of vectors \mathbf{y}_n are of course

identically 0, and we study in this paper the behaviour of $M^2 \times M^2$ matrix $\hat{\mathbf{C}}_N$ whose entries $(\hat{\mathbf{C}}_{i,k,j,l})_{1 \leq i,k,j,l \leq M}$ coincide with the traditional empirical estimate of the joint cumulant $c_4(\mathbf{y}_{i,n}, \mathbf{y}_{k,n}, \mathbf{y}_{j,n}^*, \mathbf{y}_{l,n}^*)$, i.e.

$$\hat{\mathbf{C}}_{i,k,j,l} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_{i,n} \mathbf{y}_{k,n} \mathbf{y}_{j,n}^* \mathbf{y}_{l,n}^* - \hat{\mathbf{R}}_{i,j} \hat{\mathbf{R}}_{k,l} - \hat{\mathbf{R}}_{i,l} \hat{\mathbf{R}}_{k,j}$$

where we recall that $M \times M$ matrix $\hat{\mathbf{R}}_N$ represents the empirical covariance matrix $\hat{\mathbf{R}}_N = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*$. Using Eq. (1), it is easily seen that

$$\hat{\mathbf{R}}_{i,j} \hat{\mathbf{R}}_{k,l} + \hat{\mathbf{R}}_{i,l} \hat{\mathbf{R}}_{k,j} = \left(2 \boldsymbol{\Gamma} \boldsymbol{\Gamma}^* \left(\hat{\mathbf{R}}_N \otimes \hat{\mathbf{R}}_N \right) \boldsymbol{\Gamma} \boldsymbol{\Gamma}^* \right)_{i,k,j,l}$$

Therefore, matrix $\hat{\mathbf{C}}_N$ can be written as

$$\hat{\mathbf{C}}_N = \hat{\mathbf{D}}_N - 2 \boldsymbol{\Gamma} \boldsymbol{\Gamma}^* \left(\hat{\mathbf{R}}_N \otimes \hat{\mathbf{R}}_N \right) \boldsymbol{\Gamma} \boldsymbol{\Gamma}^* \quad (2)$$

where matrix $\hat{\mathbf{D}}_N$ is defined by

$$\hat{\mathbf{D}}_N = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n \otimes \mathbf{y}_n) (\mathbf{y}_n \otimes \mathbf{y}_n)^* \quad (3)$$

We remark that for each n , $\boldsymbol{\Gamma} \boldsymbol{\Gamma}^* (\mathbf{y}_n \otimes \mathbf{y}_n)$ coincides with $\mathbf{y}_n \otimes \mathbf{y}_n$ so that the range of $\hat{\mathbf{D}}_N$ is included in the column space of $\boldsymbol{\Gamma} \boldsymbol{\Gamma}^*$. Therefore, matrix $\hat{\mathbf{C}}_N$ is rank deficient, and its range is included into the $M(M+1)/2$ -dimensional space generated by the columns of $\boldsymbol{\Gamma}$. This point will be used in section 4 below.

It is clear that if $N \rightarrow +\infty$ while M remains fixed, the law of large number implies that each element of $\hat{\mathbf{C}}_N$ converges towards 0. This implies $\|\hat{\mathbf{C}}_N\| \rightarrow 0$ because the size of matrix $\hat{\mathbf{C}}_N$ does not scale with N . When both M and N converge towards $+\infty$, the convergence of the individual entries of $\hat{\mathbf{C}}_N$ towards 0 no longer implies the convergence of $\|\hat{\mathbf{C}}_N\|$ towards 0. In the two following sections, we provide some results concerning the asymptotic behaviour of $\hat{\mathbf{C}}_N$ and $\|\hat{\mathbf{C}}_N\|$ when M and N converge to $+\infty$.

3 Conditions under which $\|\hat{\mathbf{C}}_N\| \rightarrow 0$.

In this section, we derive conditions on M and N under which $\|\hat{\mathbf{C}}_N\| \rightarrow 0$. For this, we prove the following proposition.

Proposition 1. *Assume that M and N both converge towards $+\infty$ in such a way that*

$$\frac{M^2 \log N}{N} \rightarrow 0 \quad (4)$$

Then it holds that

$$\left\| \hat{\mathbf{D}}_N - 2\sigma^4 \boldsymbol{\Gamma} \boldsymbol{\Gamma}^* \right\| \rightarrow 0 \text{ a.s.} \quad (5)$$

and that

$$\|\hat{\mathbf{C}}_N\| \rightarrow 0 \quad (6)$$

Moreover, almost surely, for N large enough, $\|\hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^*\|$ and $\|\hat{\mathbf{C}}_N\|$ are less than $\mu \left(\frac{\log N}{N}\right)^{1/2} (M + \log N)$ for some positive constant μ .

We just provide a sketch of the proof of Proposition 1. We first mention that $\mathbb{E}((\mathbf{y}_n \otimes \mathbf{y}_n)(\mathbf{y}_n \otimes \mathbf{y}_n)^*) = \mathbb{E}(\hat{\mathbf{D}}_N) = 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^*$. We now provide some insights on (5). For this, we recall the matrix Bernstein inequality (see Theorem 6.6.1 in [13]).

Theorem 1. *Let $(\mathbf{S}_n)_{n=1,\dots,N}$ be a sequence of Hermitian zero mean i.i.d. $m \times m$ random matrices satisfying $\sup_{n=1,\dots,N} \|\mathbf{S}_n\| \leq \kappa_N$ for some deterministic constant κ_N . If $\bar{\mathbf{S}}_N$ denotes matrix $\sum_{n=1}^N \mathbf{S}_n$, and if $v_N = \mathbb{E}(\bar{\mathbf{S}}_N)^2$, then, for each $\epsilon > 0$, it holds that*

$$\mathbb{P}(\|\bar{\mathbf{S}}_N\| > \epsilon) \leq m \exp\left(\frac{-\epsilon^2}{2v_N + \kappa_N \epsilon/3}\right) \quad (7)$$

We denote by \mathbf{R}_n matrix

$$\mathbf{R}_n = \frac{1}{N} ((\mathbf{y}_n \otimes \mathbf{y}_n)(\mathbf{y}_n \otimes \mathbf{y}_n)^* - \mathbb{E}(\mathbf{y}_n \otimes \mathbf{y}_n)(\mathbf{y}_n \otimes \mathbf{y}_n)^*) \quad (8)$$

We have thus to establish that if $\bar{\mathbf{R}}_N = \sum_{n=1}^N \mathbf{R}_n$, then $\|\bar{\mathbf{R}}_N\| \rightarrow 0$ almost surely. For this, we prove that if $\epsilon = \mu \left(\frac{\log N}{N}\right)^{1/2} (M + \log N)$, it exists a constant μ for which

$$\sum_{N=1}^{+\infty} \mathbb{P}(\|\bar{\mathbf{R}}_N\| > \epsilon_N) < +\infty \quad (9)$$

If (9) holds, Borel-Cantelli's lemma implies that almost surely, for N large enough, $\|\bar{\mathbf{R}}_N\| \leq \epsilon_N$, which leads to the conclusion that $\|\bar{\mathbf{R}}_N\|$ converges towards 0 at rate ϵ_N . Bernstein inequality could be used to evaluate an upper bound of $\mathbb{P}(\|\bar{\mathbf{R}}_N\| > \epsilon_N)$ if the norms of matrices $(\mathbf{R}_n)_{n=1,\dots,N}$ were bounded everywhere by a constant κ_N . As this property does not hold, it is possible to use a classical truncation argument (see e.g. [12]), and consider matrices $(\mathbf{S}_n)_{n=1,\dots,N}$ (resp. $(\mathbf{T}_n)_{n=1,\dots,N}$) defined in the same way than $(\mathbf{R}_n)_{n=1,\dots,N}$ in (8), but when \mathbf{y}_n is replaced by $\mathbf{y}_n \mathbb{1}_{(\|\mathbf{y}_n\|^2 \leq \alpha_N)}$ (resp. by $\mathbf{y}_n \mathbb{1}_{(\|\mathbf{y}_n\|^2 > \alpha_N)}$) where α_N is a well chosen deterministic constant. Due to the lack of space, we do not provide more details.

It remains to justify that (5) implies (6). For this, we recall that $\mathbf{\Gamma}^* \mathbf{\Gamma} = \mathbf{I}$, and express $\hat{\mathbf{C}}_N$ as

$$\hat{\mathbf{C}}_N = \hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^* + 2\mathbf{\Gamma} \mathbf{\Gamma}^* \left(\sigma^4 \mathbf{I} - \hat{\mathbf{R}}_N \otimes \hat{\mathbf{R}}_N\right) \mathbf{\Gamma} \mathbf{\Gamma}^* \quad (10)$$

We remark that (4) implies that $\frac{M \log N}{N} \rightarrow 0$. Using the same approach as above, it is easily seen that $\frac{M \log N}{N} \rightarrow 0$ implies that $\|\hat{\mathbf{R}}_N - \sigma^2 \mathbf{I}_M\| \rightarrow 0$, and that, almost surely, for N large enough, $\|\hat{\mathbf{R}}_N - \sigma^2 \mathbf{I}_M\|$ is less $\mu \left(\frac{\log N}{N}\right)^{1/2} (M + \log N)^{1/2}$ for some positive constant μ . Using this, it is easy to check that $\|\hat{\mathbf{R}}_N \otimes \hat{\mathbf{R}}_N - \sigma^4 \mathbf{I}\|$ converges towards 0 at the same rate that $\|\hat{\mathbf{R}}_N - \sigma^2 \mathbf{I}_M\|$. (10) and the triangular inequality completes the proof.

We now comment Proposition (1). We first remark that, up to the term $\log N$, condition (4) is optimal because, as shown below in section 4, if $\frac{M^2}{N}$ converges towards a non zero constant, then (6) does not hold. Next, it is easy to verify that Proposition 1 holds if $(\mathbf{y}_n)_{n=1, \dots, N}$ is a temporally white, but spatially correlated Gaussian noise: in this case, \mathbf{y}_n can be written as $\mathbf{y}_n = \mathbf{R}^{1/2} \mathbf{v}_n$ where \mathbf{R} represents the covariance matrix of vectors $(\mathbf{y}_n)_{n=1, \dots, N}$, and where $(\mathbf{v}_n)_{n=1, \dots, N}$ is a temporally and spatially white Gaussian noise. Using Proposition 1 to sequence $(\mathbf{v}_n)_{n=1, \dots, N}$ leads immediately to the conclusion that the estimated fourth-order cumulants matrix of $(\mathbf{y}_n)_{n=1, \dots, N}$ converges towards 0 provided the spectral norm of \mathbf{R} remains bounded when $N \rightarrow +\infty$. It would of course be useful to evaluate the behaviour of the estimated fourth-order cumulants matrix in the presence of a linear mixing of K independent signals. We feel that the use of similar tools should lead to the conclusion that $\hat{\mathbf{C}}_N$ converges towards the true fourth-order cumulants matrix provided $\frac{M^2 \log N}{N} \rightarrow 0$. This condition has to be compared with the condition $\frac{M \log N}{N} \rightarrow 0$ which ensures that the empirical covariance matrix converges towards the true covariance matrix in the spectral norm sense: estimating the true fourth-order cumulants matrix necessitates a much larger number of samples. We also conjecture that the consistent estimation of the $2p$ -order cumulant matrix would need that $\frac{M^{p/2} \log N}{N} \rightarrow 0$.

4 Study of the case where N and M^2 are of the same order of magnitude.

We now consider the asymptotic regime M, N converge towards $+\infty$ in such a way that $\frac{M^2}{N}$ converges towards a constant. In the following, we denote by L the integer $L = \frac{M(M+1)}{2}$, and define c_N as $c_N = \frac{L}{N}$. In the present asymptotic regime, c_N converges towards $c_* > 0$. In order to simplify the presentation of the following results, we assume that $c_* \leq 1$.

We first recall that the empirical eigenvalue distribution of an hermitian $m \times m$ random matrix \mathbf{B} is the random probability distribution $\frac{1}{m} \sum_{k=1}^m \delta_{\lambda_k}$ where $(\lambda_k)_{k=1, \dots, m}$ are the eigenvalues of \mathbf{B} , and where δ_λ represents the Dirac distribution at point λ . It is well known that under certain assumptions on \mathbf{B} , its empirical eigenvalue distribution converges weakly almost surely towards a deterministic probability distribution when $m \rightarrow +\infty$. Unformally, this means that the histogram of the eigenvalues of a realization of \mathbf{B} tends to accumulate around the graph of a deterministic probability distribution when m increases.

For example, if $\mathbf{B} = \frac{1}{p} \mathbf{A} \mathbf{A}^*$ where \mathbf{A} is a $m \times p$ random matrix with zero mean and variance σ^2 i.i.d. entries, when m and p both converge towards $+\infty$ in a such a way that $d_m = \frac{m}{p}$ converges towards $d_* > 0$, the empirical eigenvalue distribution of \mathbf{B} converges towards the so-called Marcenko-Pastur distribution with parameter (σ^2, d_*) denoted $\text{MP}(\sigma^2, d_*)$ in the following (see e.g. [10] or the tutorial [4]). If $d_* \leq 1$, distribution $\text{MP}(\sigma^2, d_*)$ is absolutely continuous, and is supported by interval $[\sigma^2(1-\sqrt{d_*})^2, \sigma^2(1+\sqrt{d_*})^2]$. It is important to remark that in the context of this particular random matrix model, the asymptotic behaviour of the measure $\frac{1}{m} \sum_{k=1}^m \delta_{\lambda_k}$ does not depend on the probability distribution of the entries of \mathbf{A} .

In the regime considered in this section, we will see that the norm of $\hat{\mathbf{C}}_N$ does not converge towards 0. Therefore, the eigenvalues of $\hat{\mathbf{C}}_N$ do not concentrate around 0, and we propose to evaluate the behaviour of the empirical eigenvalue distribution of $\hat{\mathbf{C}}_N$ to understand how $\hat{\mathbf{C}}_N$ deviates from the null matrix. For this, we notice that in the present asymptotic regime, it holds that $\frac{M \log N}{N} \rightarrow 0$. Therefore, matrix $\hat{\mathbf{R}}_N$ converges towards $\sigma^2 \mathbf{I}_M$ in the spectral norm sense. Consequently, matrix $\hat{\mathbf{C}}_N$ has the same behaviour than matrix $\hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^*$ in the sense that

$$\left\| \hat{\mathbf{C}}_N - (\hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^*) \right\| \rightarrow 0 \text{ a.s.} \quad (11)$$

Therefore, the eigenvalues of $\hat{\mathbf{C}}_N$ behave like the eigenvalues of $\hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^*$. We thus study the empirical eigenvalue distribution of the latter matrix. For this, we recall that the range of $\hat{\mathbf{D}}_N$ coincides with the range of $\mathbf{\Gamma}$. Hence, it holds that

$$\mathbf{\Gamma} \mathbf{\Gamma}^* \hat{\mathbf{D}}_N \mathbf{\Gamma} \mathbf{\Gamma}^* = \hat{\mathbf{D}}_N$$

and matrix $\hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^*$ can be written as

$$\hat{\mathbf{D}}_N - 2\sigma^4 \mathbf{\Gamma} \mathbf{\Gamma}^* = \mathbf{\Gamma} (\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - 2\sigma^4 \mathbf{I}_L) \mathbf{\Gamma}^* \quad (12)$$

where $\mathbf{\Sigma}_N$ is the $L \times N$ random matrix defined by

$$\mathbf{\Sigma}_N = \frac{1}{\sqrt{N}} \mathbf{\Gamma}^* (\mathbf{y}_1 \otimes \mathbf{y}_1, \dots, \mathbf{y}_N \otimes \mathbf{y}_N)$$

The eigenvalues of $\hat{\mathbf{C}}_N$ are thus 0 with multiplicity $M^2 - L = M(M-1)/2$ as well as the eigenvalues of matrix $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - 2\sigma^4 \mathbf{I}_L$. In order to evaluate the asymptotic behaviour of the eigenvalue distribution of $\hat{\mathbf{C}}_N$, it is thus sufficient to study the empirical eigenvalue distribution of matrix $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$. We denote by ξ_1, \dots, ξ_N the columns of $\mathbf{\Sigma}_N$. It is clear that vectors $(\xi_n)_{n=1, \dots, N}$ are independent and identically distributed, that $\mathbb{E}(\xi_n \xi_n^*) = \frac{2\sigma^4}{N} \mathbf{I}_L$, but that for each n , the entries of vector ξ_n are of course not independent. However, the behaviour of the eigenvalues of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ behave as if the entries of $\mathbf{\Sigma}_N$ were i.i.d. More precisely, the following result holds.

Theorem 2. *The empirical eigenvalue distribution of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ converges almost surely towards $\text{MP}(2\sigma^4, c_*)$. Moreover, the largest eigenvalue and the smallest*

eigenvalue of $\Sigma_N \Sigma_N^*$ converge almost surely towards $2\sigma^4(1 + \sqrt{c_*})^2$ and $2\sigma^4(1 - \sqrt{c_*})^2$ respectively, and this implies that for each $\epsilon > 0$, almost surely, all the eigenvalues of $\Sigma_N \Sigma_N^*$ lie in the interval $[2\sigma^4(1 - \sqrt{c_N})^2 - \epsilon, 2\sigma^4(1 + \sqrt{c_N})^2 + \epsilon]$ for N large enough. Finally, 0 is eigenvalue of $\hat{\mathbf{C}}_N$ with multiplicity $M^2 - L = M(M - 1)/2$. For each $\epsilon > 0$, almost surely, the L remaining eigenvalues of $\hat{\mathbf{C}}_N$ are located in the interval $[-2\sigma^4(2\sqrt{c_N} - c_N) - \epsilon, 2\sigma^4(2\sqrt{c_N} + c_N) + \epsilon]$ for N large enough, and the distribution of these remaining eigenvalues converge towards a translated version of $\text{MP}(2\sigma^4, c_*)$.

The convergence of the eigenvalue distribution of $\Sigma_N \Sigma_N^*$ towards $\text{MP}(2\sigma^4, c_*)$ follows immediately from the general results of [9]. The convergence of the extreme eigenvalues of $\Sigma_N \Sigma_N^*$ is more demanding, and follows an approach developed in a different context in [5] and [11]. The behaviour of the eigenvalues of $\hat{\mathbf{C}}_N$ is a direct consequence of (11) and (12).

Theorem 2 implies that the non zero eigenvalues of $\hat{\mathbf{C}}_N$ lie in the neighbourhood of an interval whose length δ_N is equal to $\delta_N = 8\sigma^4\sqrt{c_N}$. δ_N is thus proportional to $2\sigma^4$ which corresponds to the fourth-order moment of the components of vectors $(\mathbf{y}_n)_{n=1, \dots, N}$. δ_N also depends on M and N through $\sqrt{c_N} = \left(\frac{M(M+1)}{2N}\right)^{1/2}$ which is nearly equal to $\frac{1}{\sqrt{2}}\left(\frac{M^2}{N}\right)^{1/2}$. The spreading of the eigenvalues thus depends on the ratio $\frac{M^2}{N}$, which, in order to be a small factor, needs N to be very large. We notice that the spreading of the eigenvalues of the empirical matrix $\hat{\mathbf{R}}_N$ is equal to $4\sigma^2\left(\frac{M}{N}\right)^{1/2}$. This tends to indicate that to estimate \mathbf{C}_N with the same accuracy than the covariance matrix \mathbf{R}_N , the sample size should be increased by a factor M .

We finally illustrate Theorem 2. In our numerical experiments, $\sigma^2 = 4$, $N = 6000$ and $M = 50$ so that $c_N = 0.21$. In Figure 1, we represent the histogram of the eigenvalues of a realization of matrix $\Sigma_N \Sigma_N^*$. It appears that the histogram fits quite well with the graph of the probability density of $\text{MP}(32, 0.21)$, and that all the eigenvalues lie in the support of the MP distribution, thus confirming the practical reliability of the first statements of Theorem 2. In Figure 2, we represent the histogram of the non zero eigenvalues of $\hat{\mathbf{C}}_N$. This time, we can observe a larger gap between the histogram and the limit distribution. This extra gap follows from the errors due to the approximation $\hat{\mathbf{R}}_N \simeq \sigma^2 \mathbf{I}_M$.

5 Concluding remarks.

In this paper, we have shown that the estimated fourth-order cumulant matrix $\hat{\mathbf{C}}_N$ of a temporally and spatially white Gaussian noise converges towards 0 in the spectral norm sense if $\frac{M^2}{N} \log N$ converges towards 0. When $\frac{M^2}{N}$ converges towards a non zero constant, the empirical eigenvalue distribution of $\hat{\mathbf{C}}_N$ converges towards a translated Marcenko-Pastur distribution, and all the eigenvalues of $\hat{\mathbf{C}}_N$ lie for N large enough in an interval whose length is a $\mathcal{O}\left(\sigma^4\left(\frac{M^2}{N}\right)^{1/2}\right)$ term. In the presence of sources, this suggests that in order to estimate accurately $\hat{\mathbf{C}}_N$

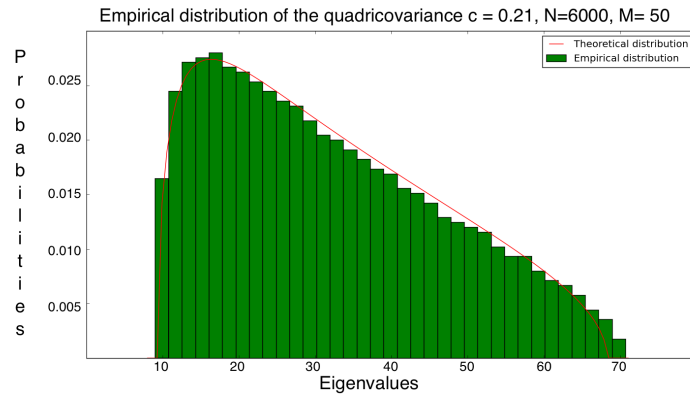


Fig. 1. Histogram of the eigenvalues of $\Sigma_N \Sigma_N^*$

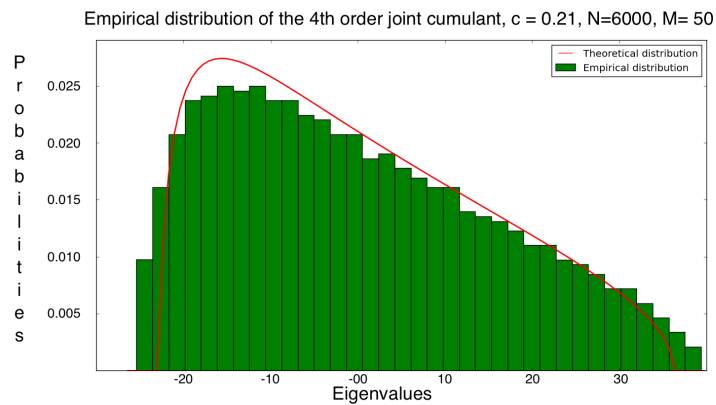


Fig. 2. Histogram of the non zero eigenvalues of \hat{C}_N

when M is large, the number of observations N should be much larger than M^2 . In the next future, we will study the behaviour of the largest eigenvalues and corresponding eigenvectors of \hat{C}_N in the presence of sources when $\frac{M^2}{N}$ does not

converge towards 0. Hopefully, this will allow to propose improved performance estimation algorithms of the mixing matrix when $\frac{M^2}{N}$ is not small enough.

References

1. L. Albéra, A. Ferréol, P. Chevalier, P. Comon, "ICAR, a tool for blind source separation using fourth order statistics only", *IEEE Transactions On Signal Processing*, vol. 53, n. 10, part 1, pp. 3633-3643, October 2005.
2. P. Bianchi, M. Debbah, M. Maëda and J. Najim, "Performance of Statistical Tests for Single Source Detection using Random Matrix Theory", *IEEE Trans. Inf. Theory*, vol. 57, no.4, pp. 2400-2419, april 2011
3. P.J. Bickel, E. Levina, "Regularized estimation of large covariance matrices", *Ann. of Statistics*, 2008, vol. 36, 199-227.
4. R. Couillet, M. Debbah, "Signal Processing in Large Systems: a New Paradigm", *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 24-39, 2013.
5. U. Haagerup, S. Thorbjornsen, "A new application of random matrices: $\text{Ext}(C_{red}^*(F_2))$ is not a group", *Annals of Mathematics*, vol. 162, no. 2, 2005.
6. S. Hiltunen, P. Loubaton, P. Chevalier, "Large system analysis of a GLRT for detection with large sensor arrays in temporally white noise", *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5409 - 5423, October 15 2015
7. S. Kritchman, 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.
8. R.R. Nadakuditi, J.W Silverstein, "Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples", *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 468 -480, June 2010.
9. A. Pajor, L.A. Pastur, "On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution", *Studia Math.* 195 (2009), 11-29.
10. L.A. Pastur, M. Shcherbina, "Eigenvalue Distribution of Large Random Matrices", *Mathematical Surveys and Monographs*, Providence: American Mathematical Society, 2011.
11. H. Schultz, "Non commutative polynomials of independent Gaussian random matrices", *Probab. Theory Relat. Fields* 131, 261-309 (2005)
12. T. Tao, "Topics in Random Matrix Theory", *Graduate Studies in Mathematics*, vol. 132, American Mathematical Society, Providence, Rhode Island, 2012.
13. J. Tropp, "An Introduction to Matrix Concentration Inequalities", *Found. Trends Mach. Learning*, Vol. 8, num. 1-2, pp. 1-230, May 2015.
14. P. Vallet, X. Mestre, P. Loubaton, "Performance analysis of an improved MUSIC DoA estimator", *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6407-6422, December 1 2015