



On the non-detectability of spiked large random tensors

Antoine Chevreuil, Philippe Loubaton

► To cite this version:

Antoine Chevreuil, Philippe Loubaton. On the non-detectability of spiked large random tensors . SSP, Jun 2018, Freiburg, Germany. hal-01798509

HAL Id: hal-01798509

<https://hal.science/hal-01798509>

Submitted on 23 May 2018

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 NON-DETECTABILITY OF SPIKED LARGE RANDOM TENSORS

A. Chevreuil and P. Loubaton

Laboratoire d'Informatique Gaspard Monge (CNRS, Université Paris-Est/MLV)
5 Bd. Descartes 77454 Marne-la-Vallée (France)

ABSTRACT

This paper addresses the detection of a low rank high-dimensional tensor corrupted by an additive complex Gaussian noise. In the asymptotic regime where all the dimensions of the tensor converge towards $+\infty$ at the same rate, existing results devoted to rank 1 tensors are extended. It is proved that if a certain parameter depending explicitly on the low rank tensor is below a threshold, then the null hypothesis and the presence of the low rank tensor are undistinguishable hypotheses in the sense that no test performs better than a random choice.

1. INTRODUCTION

The problem of testing whether an observed $n_1 \times n_2$ matrix $\mathbf{Y}$ is either a zero-mean independent identically distributed Gaussian random matrix $\mathbf{Z}$ with variance $\frac{1}{n_2}$, or $\mathbf{X}_0 + \mathbf{Z}$ (where $\mathbf{X}_0$ is a low rank matrix: a useful signal, called also *spike*) is a fundamental problem arising in numerous applications such as the detection of low-rank multivariate signals or the Gaussian hidden clique problem. When the two dimensions n_1, n_2 converge towards ∞ at the same rate, the rank of $\mathbf{X}_0$ remaining fixed, the context is this of the so-called additive spiked large random matrix models. Various results on the singular values of $\mathbf{X}_0 + \mathbf{Z}$ have been established; in particular it is possible to show that the Generalized Likelihood Ratio Test (GLRT) is consistent (*i.e.* the probability of false alarm and the probability of missed detection both converge towards 0 when n_1, n_2 converge towards $+\infty$ in such a way that $n_1/n_2 \rightarrow c > 0$) if and only the largest singular value of $\mathbf{X}_0$ is above the threshold $c^{1/4}$ (see *e.g.* [12], [3], [2]).

In a number of real life problems, the observation is not a matrix, but a tensor $\mathbf{Y}$ of order $d \geq 3$, *i.e.* a d -dimensional array $\mathbf{Y} = \mathbf{Y}_{i_1, i_2, \dots, i_d}$ where for each $k = 1, \dots, d$, $i_k \in [1, \dots, n_k]$. In this context, the generalization of the above matrix hypothesis testing problem becomes: test that the observed order $d \geq 3$ tensor is either a zero-mean independent identically distributed Gaussian random tensor $\mathbf{Z}$, or the sum of $\mathbf{Z}$ and a low rank deterministic tensor

$$\mathbf{X}_0 = \sum_{i=1}^r \lambda_i \mathbf{x}_0^{(1,i)} \otimes \mathbf{x}_0^{(2,i)} \otimes \dots \otimes \mathbf{x}_0^{(d,i)} \quad (1)$$

where r is called the rank of $\mathbf{X}_0$. Here $(\lambda_i)_{i=1, \dots, r}$ are strictly positive real numbers, and for each $i = 1, \dots, r$ and $k = 1, \dots, d$, $\mathbf{x}_0^{(k,i)}$ is a $n_k \times 1$ unit norm vector. Recent works (see *e.g.* [8, 11, 10, 14]) addressed the detection/estimation of $\mathbf{X}_0$ when r is reduced to 1 and when the dimensions $n_1, \dots, n_d$ converge towards ∞ at the same rate. We also mention that [8] and [14] only considered the case where the rank 1 tensor $\mathbf{X}_0$ is symmetric, that is: $n_1 = n_2 =$

$\dots = n_d$ and the d vectors $(\mathbf{x}_0^{(k,i)})_{i=1, \dots, r}$ are equal. Since the concept of singular value decomposition cannot be extended to tensors, *ad hoc* statistical strategies have been considered to prove the (non)-existence of consistent tests: [11] and [14] ($r = 1$) established that if λ_1 is larger than a certain upper bound, then consistent detection of $\mathbf{X}_0$ is possible. In the other direction, [10][14] (again, $r = 1$) proved that if λ_1 is less than a certain lower bound (which is strictly less than the above upper bound), then $\mathbf{X}_0$ is non-detectable in the sense that any test behaves as a random choice between the two hypotheses. This is a remarkable phenomenon because such a behaviour is not observed in the matrix case ($d = 2$): indeed, in this context, if the largest eigenvalue of $\mathbf{X}_0$ is below $c^{1/4}$, it is proved in [13] ($r = 1$) that there exist statistical tests having a better performance than a random choice - a result that [10][14] obtained a different way.

The replica method has been successfully considered [4][9]. In these contributions, the model does not match exactly ours since 1) the spike is assumed symmetric, *i.e.* $\mathbf{X}_0 = \sum_{i=1}^r \lambda_i (\mathbf{x}_0^{(i)})^{\otimes d}$ and 2) the rows of the matrix $(\mathbf{x}_0^{(1)}, \dots, \mathbf{x}_0^{(r)})$ are random i.i.d. with a known distribution (the prior). When $r = 1$, and the prior is of the Rademacher type, the observed tensor follows the pure p-spin model [15]: in an illuminating contribution [4], a tight threshold when $d \geq 3$ is provided (above which consistent detection is possible and under which any detector performs as a random guess). The case $r \geq 1$ with a general prior is addressed in [9]: there, the estimation of the spike is considered rather than the detection; specifically, the asymptotic performance of the MMSE is computed and an estimation threshold is deduced. This latter is rigorously proved when $r = 1$. The threshold is not explicit and intrinsically depends on the prior.

In the present contribution, we follow the methodology of [11][10][14] and extend it to the general case $r \geq 1$: though suboptimal (the thresholds provided are not tight in general), the machinery is much lighter than this of the replica method, it provides explicit bounds for the non-detectability and lastly allows one to deal with deterministic spikes. Precisely, we find out a simple sufficient condition on the spike $\mathbf{X}_0$ under which $\mathbf{X}_0$ is non-detectable. The problem of finding conditions under which the existence of a consistent detection is guaranteed is not addressed here.

2. MODEL, NOTATION, AND BACKGROUND

The order- d tensors are complex-valued, and it is assumed that $n_1 = n_2 = \dots = n$ in order to simplify the notations. The set $\otimes^d \mathbb{C}^n$ is a complex vector-space endowed with the standard scalar product

$$\forall \mathbf{X}, \mathbf{Y} \in \otimes^d \mathbb{C}^n \quad \langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i_1, \dots, i_d} \mathbf{X}_{i_1, \dots, i_d} \bar{\mathbf{Y}}_{i_1, \dots, i_d}$$

and the Frobenius norm $\|\mathbf{X}\|_F = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle}$.

We are grateful to Lenka Zdeborová for bringing to our knowledge important bibliographical references.

The spike ("the signal") is assumed to be a tensor of fixed rank r following (1). Along this contribution, n is large or, mathematically, $n \rightarrow \infty$. We hence have for each n a set of $n \times 1$ vectors $(\mathbf{x}_0^{(k,i)})_{k=1,\dots,d,i=1,\dots,r}$. For each $k = 1, \dots, d$, we denote by $\chi_0^{(k)}$ the $n \times r$ matrix $\chi_0^{(k)} = (\mathbf{x}_0^{(k,1)}, \dots, \mathbf{x}_0^{(k,r)})$. We impose a non-erratic asymptotic behavior of the spike, and specifically, as all the vectors $\mathbf{x}_0^{(k,i)} \in \mathbb{C}^{n \times 1}$ have unit norm, we suppose that for all i, j , $\langle \mathbf{x}_0^{(k,i)}, \mathbf{x}_0^{(k,j)} \rangle = (\chi_0^{(k)*} \chi_0^{(k)})_{i,j}$ converges as $n \rightarrow \infty$. The rate of convergence is a technical aspect that is out of the scope of this contribution: we will simply assume that the matrices $(\chi_0^{(k)*} \chi_0^{(k)})_{k=1,\dots,d}$ do not depend on n . We define the SVD of $\chi_0^{(k)}$ as $\mathbf{U}_k \begin{pmatrix} \Sigma_k & 0 \end{pmatrix} \mathbf{V}_k^*$ for $\mathbf{U}_k$ and $\mathbf{V}_k$ unitary matrices respectively of size $n \times n$ and $r \times r$ and Σ_k a diagonal matrix with non-negative entries on the diagonal. $\mathbf{V}_k$ and Σ_k do not depend on n because $\chi_0^{(k)*} \chi_0^{(k)} = \mathbf{V}_k \Sigma_k^2 \mathbf{V}_k^*$.

We denote by $\mathbf{Z}$ the noise tensor, and assume that its entries are $\mathcal{N}_{\mathbb{C}}(0, 1/n)$ independent identically distributed complex circular Gaussian random variables.

In the following, we consider the alternative $\mathcal{H}_0 : \mathbf{Y} = \mathbf{Z}$ versus $\mathcal{H}_1 : \mathbf{Y} = \mathbf{X}_0 + \mathbf{Z}$. We denote by $p_{1,n}(\mathbf{y})$ the probability density of $\mathbf{Y}$ under $\mathcal{H}_1$ and $p_{0,n}(\mathbf{y})$ the density of $\mathbf{Y}$ under $\mathcal{H}_0$. $\Lambda(\mathbf{Y}) = \frac{p_1(\mathbf{Y})}{p_0(\mathbf{Y})}$ is the likelihood ratio and we denote by $\mathbb{E}_0$ the expectation under $\mathcal{H}_0$. We now recall the fundamental information geometry results used in [10] in order to address the detection problem. The following properties are well known (see also [1] section 3):

- (i) If $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2]$ is bounded, then no consistent detection test exists.
- (ii) If moreover $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2] = 1 + o(1)$, then the total variation distance between $p_{0,n}$ and $p_{1,n}$ converges towards 0, and no test performs better than a decision at random.

Therefore, the computation of the second order moment of $\Lambda(\mathbf{Y})$ under $p_{0,n}$ may provide insights on the detection. We however notice that conditions (i) and (ii) are only sufficient. In particular, if $\limsup_n \mathbb{E}_0 [\Lambda(\mathbf{Y})^2] = +\infty$, nothing can be inferred on the behaviour of the detection problem when $n \rightarrow +\infty$.

3. EXPRESSION OF THE SECOND-ORDER MOMENT.

The density of $\mathbf{Z}$, seen as a collection of n^d complex-valued random variables, is obviously $p_{0,n}(\mathbf{z}) = \kappa_n \exp(-n \|\mathbf{z}\|_F^2)$ where $\kappa_n = (\frac{n}{\pi})^{n^d}$. On the one hand, we notice that the second-order moment approach is not suited to the deterministic model of the spike as presented previously. Indeed, in this case $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2]$ has the simple expression $\exp(2n \|\mathbf{X}_0\|_F^2)$ and always diverges. On the other hand, the noise tensor shows an invariance property: if $\Theta_1, \dots, \Theta_d$ are unitary $n \times n$ matrices, then the density of the mode products $(\Theta_1 \otimes \Theta_2 \dots \otimes \Theta_d) \mathbf{Z}$ equals this of $\mathbf{Z}$. For $d = 2$, the notation $(\Theta_1 \otimes \Theta_2) \mathbf{Z}$ simply means $\Theta_1 \mathbf{Z} \Theta_2$ and for a general d , $((\Theta_1 \otimes \Theta_2 \dots \otimes \Theta_d) \mathbf{Z})_{i_1, \dots, i_d}$ is

$$\sum_{\ell_1, \dots, \ell_d} (\Theta_1)_{i_1, \ell_1} (\Theta_2)_{i_2, \ell_2} \dots (\Theta_d)_{i_d, \ell_d} \mathbf{Z}_{\ell_1, \dots, \ell_d}.$$

We hence modify the data according to the procedure: we pick i.i.d. complex Haar samples $\Theta_1, \dots, \Theta_d$ and consider the "new" data tensor defined as $(\Theta_1 \otimes \Theta_2 \dots \otimes \Theta_d) \mathbf{Y}$. This does not affect the distribution of the noise, but this amounts to assume an artificial prior

on the spike. Indeed, the vectors $\mathbf{x}_0^{(k,i)}$ are replaced by $\Theta_k \mathbf{x}_0^{(k,i)}$. They are all uniformly distributed on the unit sphere of $\mathbb{C}^n$ and for $k \neq l$, vectors $\Theta_k \mathbf{x}_0^{(k,i)}$ and $\Theta_l \mathbf{x}_0^{(l,j)}$ are independent for each i, j . However, vectors $(\Theta_k \mathbf{x}_0^{(k,i)})_{i=1,\dots,r}$ are not independent. In the following, the data and the noise tensors after this procedure are still denoted respectively by $\mathbf{Y}$ and $\mathbf{Z}$. This transformation of the spike is an extension of a trick used in Section III.C of [10].

We are now in position to give a closed-form expression of the second-order moment of $\Lambda(\mathbf{Y})$. We have $p_{1,n}(\mathbf{Y}) = \mathbb{E}_X [p_{0,n}(\mathbf{Y} - \mathbf{X})]$ where $\mathbb{E}_X$ is the mathematical expectation over the distribution of the spike, or equivalently over the Haar matrices $(\Theta_k)_{k=1,\dots,d}$. It holds that

$$\mathbb{E}_0 [\Lambda(\mathbf{Y})^2] = \mathbb{E}_{X, X'} [\exp(2n \Re \langle \mathbf{X}, \mathbf{X}' \rangle)] = \mathbb{E}_{X, X'} \left[\exp \left(2n \Re \sum_{i,j=1}^r \lambda_i \lambda_j \prod_{k=1}^d \langle (\Theta_k')^* \Theta_k \mathbf{x}_0^{(k,i)}, \mathbf{x}_0^{(k,j)} \rangle \right) \right]$$

where $\mathbb{E}_{X, X'}$ is over independent copies $\mathbf{X}, \mathbf{X}'$ of the spike associated respectively with $(\Theta_k)_{k=1,\dots,d}$ and $(\Theta_k')_{k=1,\dots,d}$. $\Re$ stands for the real part. As Θ_k and Θ_k' are Haar and independent, then $(\Theta_k')^* \Theta_k$ is also Haar distributed and $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2] = \mathbb{E} [\exp(2n \eta)]$, where the expectation is over the i.i.d. Haar matrices $\Theta_1, \Theta_2, \dots, \Theta_d$ and

$$\eta = \Re \sum_{i,j=1}^r \lambda_i \lambda_j \prod_{k=1}^d \underbrace{\langle \Theta_k \mathbf{x}_0^{(k,i)}, \mathbf{x}_0^{(k,j)} \rangle}_{\xi_k^{(i,j)}}. \quad (2)$$

η may be factored as $\eta = \Re \left[\lambda^T \left(\odot_{k=1}^d \left(\chi_0^{(k)*} \Theta_k \chi_0^{(k)} \right) \right) \lambda \right]$. In the latter equation, $\odot$ stands for the Hadamard product of matrices. The ultimate simplification comes from the SVD of $\chi_0^{(k)}$:

$$\chi_0^{(k)*} \Theta_k \chi_0^{(k)} = \mathbf{V}_k \begin{pmatrix} \Sigma_k & 0 \end{pmatrix} \mathbf{U}_k^* \Theta_k \mathbf{U}_k \begin{pmatrix} \Sigma_k \\ 0 \end{pmatrix} \Sigma_k \mathbf{V}_k^*.$$

Firstly, $\mathbf{U}_k^* \Theta_k \mathbf{U}_k$ has the same distribution as Θ_k ; secondly, we may associate with any Θ_k its upper $r \times r$ block, that we will denote Ψ_k . As a conclusion, we may express η as

$$\eta = \Re \left[\lambda^T \left(\odot_{k=1}^d (\mathbf{V}_k \Sigma_k \Psi_k \Sigma_k^* \mathbf{V}_k^*) \right) \lambda \right]. \quad (3)$$

4. EXTENDING KNOWN RESULTS

When $r = 1$, Montanari *et al.* [10] found a bound on the parameter λ_1 ensuring that $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2]$ is bounded. In this case, η has a simple expression since $\eta = \lambda^2 \Re \prod_{k=1}^d \xi_k$ where the $(\xi_k)_{k=1,\dots,d}$ are i.i.d. distributed as the first component of a uniform vector of the unit sphere of $\mathbb{C}^n$. As in [10], we introduce

$$\beta_d^{2nd} = \sqrt{\min_{u \in [0,1]} -\frac{1}{u^d} \log(1 - u^2)}. \quad (4)$$

Adapting the result of the aforementioned article to the complex-circular context is straight-forward:

Theorem 1 (case $r=1$ (Montanari *et al.*)). *Let $\xi_1, \dots, \xi_d$ be i.i.d. distributed as the first component of a vector uniformly distributed on the unit sphere of $\mathbb{C}^n$.*

If $\lambda_1 < \sqrt{\frac{d}{2}} \beta_d^{2nd}$ then $\mathbb{E}_0 [\exp(2n \lambda_1^2 \Re \prod_{k=1}^d \xi_k)]$ is bounded; moreover, if $d > 2$, the above expectation is $1 + o(1)$.

This non-obvious result may be used in order to derive a condition ensuring that hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are indistinguishable when $r > 1$. In this respect, recall the expansion (2). Thanks to the Hölder inequality, $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2]$ is upper bounded by (see (2) for the definition of $\xi_k^{(i,j)}$)

$$\prod_{i,j=1}^r \mathbb{E}^{1/p_{i,j}} \left[\exp \left(2np_{i,j} \lambda_i \lambda_j \Re \prod_{k=1}^d \xi_k^{(i,j)} \right) \right] \quad (5)$$

for any non-negative numbers $p_{i,j}$ such that $\sum_{i,j} \frac{1}{p_{i,j}} = 1$. For fixed i, j , we notice that the random variables $(\xi_k^{(i,j)})_{k=1,\dots,d}$ verify the condition of Theorem 1. Any of the expectations in (5) are hence upper-bounded when $n \rightarrow \infty$ provided that, for all i, j : $p_{i,j} \lambda_i \lambda_j < \frac{d}{2} (\beta_d^{2nd})^2$. Choosing eventually $p_{i,j} = \frac{(\sum_p \lambda_p)^2}{\lambda_i \lambda_j}$, we deduce

Theorem 2 (case $r \geq 1$ - extension of Theorem 1). *If $\sum_{i=1}^r \lambda_i < \sqrt{\frac{d}{2} \beta_d^{2nd}}$ then $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2]$ is bounded. If moreover $d > 2$, we have $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2] = 1 + o(1)$ and the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are indistinguishable.*

Remark 3. Due to the use of the Hölder inequality, Theorem 2 is suboptimum in general. The inequality is patently an equality when $\forall k, i, j, \mathbf{x}_0^{(k,i)} = \mathbf{x}_0^{(k,j)}$, i.e. the spike has rank $r = 1$ and amplitude $\sum_{i=1}^r \lambda_i$.

5. A TIGHTER BOUND

The main result of our contribution is the following

Theorem 4 (case $r \geq 1$). *We define $\eta_{\max}$ as*

$$\eta_{\max} = \boldsymbol{\lambda} \left(\odot_{k=1}^d \boldsymbol{\chi}_0^{(k)*} \boldsymbol{\chi}_0^{(k)} \right) \boldsymbol{\lambda}. \quad (6)$$

If $\sqrt{\eta_{\max}} < \sqrt{\frac{d}{2} \beta_d^{2nd}}$ then, for $d > 2$, $\mathbb{E}_0 [\Lambda(\mathbf{Y})^2] = 1 + o(1)$.

Before providing elements of the proof of the above result, we may briefly justify why the bound in Theorem 4 is tighter than this of Theorem 2, whatever the choice of $\boldsymbol{\lambda}$. On the one hand, indeed, $(\sum_i \lambda_i)^2 = \boldsymbol{\lambda}^T \mathbf{J} \boldsymbol{\lambda}$ where $\mathbf{J}$ is the $r \times r$ matrix having all its entries equal to 1. On the other hand, all the vectors $\mathbf{x}_0^{(k,i)}$ are normalized and consequently, any of the diagonal entries of $\boldsymbol{\chi}_0^{(k)*} \boldsymbol{\chi}_0^{(k)}$ equals 1 and for any $i \neq j$, $|\langle \boldsymbol{\chi}_0^{(k)*} \boldsymbol{\chi}_0^{(k)} \rangle_{i,j}| \leq 1$. This proves that $(\sum_i \lambda_i)^2 - \eta_{\max} = \boldsymbol{\lambda}^T (\mathbf{J} - \odot_{k=1}^d \boldsymbol{\chi}_0^{(k)*} \boldsymbol{\chi}_0^{(k)}) \boldsymbol{\lambda} \geq 0$.

We provide the key elements of the proof of Theorem 4. Remind that we are looking for a condition on the spike under which $\mathbb{E}[\exp(2n\eta)]$ is bounded. Evidently, the divergence may occur only when $\eta > 0$. We hence consider $E_1 = \mathbb{E}[\exp(2n\eta) \mathbf{1}_{\eta > \epsilon}]$ and $E_2 = \mathbb{E}[\exp(2n\eta) \mathbf{1}_{\eta \leq \epsilon}]$, and prove that under the condition $\sqrt{\eta_{\max}} < \sqrt{\frac{d}{2} \beta_d^{2nd}}$, for a certain small enough ϵ , $E_1 = o(1)$ (for $d \geq 2$) and that $E_2 = 1 + o(1)$ (for $d > 2$).

The E_1 term. It is clear that the boundedness of the integral E_1 is achieved when η rarely deviates from 0. As remarked in [10], the natural machinery to consider to understand E_1 is this of the Large Deviation Principle (LDP). In essence, if η follows the LDP with rate n , there can be found a certain non-negative function called Good Rate Function (GRF) I_η such that for any Borel set A of $\mathbb{R}$,

$\frac{1}{n} \log \mathbb{P}(\eta \in A)$ converges towards $\sup_{x \in A} -I_\eta(x)$. The existence of a GRF allows one to analyze the asymptotic behaviour of integral E_1 . Indeed, the Varadhan lemma (see Theorem 4.3.1 in [6]) states that $\frac{1}{n} \log \mathbb{E}[\exp(2n\eta) \mathbf{1}_{\eta > \epsilon}] \rightarrow \sup_{x > \epsilon} (2x - I_\eta(x))$ and hence the E_1 term converges towards 0 when $\sup_{x > \epsilon} (2x - I_\eta(x)) < 0$.

We thus justify that η follows a Large Deviation Principle with rate n , and we compute a lower bound of its GRF. For this, we use that for each k , the random matrix Ψ_k defined in (3) follows a LDP with rate n and that its GRF at the parameter $\boldsymbol{\psi} \in \mathbb{C}^{r \times r}$ (we may evidently take $\|\boldsymbol{\psi}\|_2 \leq 1$) is $\log \det(\mathbf{I}_r - \boldsymbol{\psi}^* \boldsymbol{\psi})$ (see Theorem 3-6 in [7]). η is a function of the i.i.d. matrices $(\Psi_k)_{k=1,\dots,d}$. Therefore, the contraction principle (see Theorem 4.2.1 in [6]) ensures that η follows a LDP with rate n and GRF I_η given, for each real x in the range of η , as the solution of the optimization problem:

$$\max_{\forall k, 0 \leq \alpha_k \leq 1} \max_{\forall k, \|\boldsymbol{\psi}_k\| = \alpha_k} \sum_{k=1}^d \log \det(\mathbf{I}_r - \boldsymbol{\psi}_k^* \boldsymbol{\psi}_k) \quad (7)$$

$$\eta(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_d) = x$$

When $d \geq 3$, the solution of this optimization problem cannot apparently be expressed in closed form. We thus just provide a lower bound of $I_\eta(x)$. When $d = 2$, it is possible to evaluate $I_\eta(x)$: see [5].

Proposition 5. *For each $x \in \mathbb{R}$, it holds that*

$$I_\eta(x) \geq -d \log \left(1 - \left(\frac{|x|}{\eta_{\max}} \right)^{2/d} \right). \quad (8)$$

where the right-hand side should be understood as $+\infty$ if $|x| \geq \eta_{\max}$.

In order to establish Proposition 5, we use the following algebraic result whose proof is omitted.

Lemma 6. *For any matrices $(\mathbf{A}_k)_{k=1,\dots,d} \in \mathbb{C}^{r \times r}$ and vector $\boldsymbol{\lambda} \in \mathbb{R}^r$, the supremum of $|\boldsymbol{\lambda}^T \odot_{k=1}^d (\mathbf{A}_k \boldsymbol{\psi}_k \mathbf{A}_k^*) \boldsymbol{\lambda}|$ over $r \times r$ matrices $\boldsymbol{\psi}_k$ such that for all k : $\|\boldsymbol{\psi}_k\|_2 = \alpha_k$ is*

$$\left(\prod_{k=1}^d \alpha_k \right) \boldsymbol{\lambda}^T \left(\odot_{k=1}^d (\mathbf{A}_k \mathbf{A}_k^*) \right) \boldsymbol{\lambda}.$$

The immediate consequence of this lemma is that the random variable η is bounded and $|\eta| \leq \eta_{\max}$ where $\eta_{\max}$ is given by (6). Moreover, take a set of matrices $\boldsymbol{\psi}_k$ such that $\|\boldsymbol{\psi}_k\|_2 = \alpha_k \in [0, 1]$; then by Lemma 6, $|\eta(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_d)| \leq (\prod_k \alpha_k) \eta_{\max}$ hence the optimization (7) is to be carried out only on the set of matrices $\boldsymbol{\psi}_k$ such that $\prod_k \alpha_k \geq \frac{|x|}{\eta_{\max}}$. On the other hand, one may use the generous bound $\log \det(\mathbf{I}_r - \boldsymbol{\psi}_k^* \boldsymbol{\psi}_k) \leq \log(1 - \|\boldsymbol{\psi}_k\|_2^2)$ and finally prove that

$$-I_\eta(x) \leq \max_{\prod_k \alpha_k \geq \frac{|x|}{\eta_{\max}}} \sum_{k=1}^d \log(1 - \alpha_k^2).$$

The supremum of the r.h.s. of this equation is achieved for balanced α_k and we immediately obtain (8). This completes the proof of Proposition 5.

Using Proposition 5 and setting $u = \left(\frac{|x|}{\eta_{\max}} \right)^{1/d}$, we obtain immediately that for each $\delta > 0$:

$$\frac{1}{n} \log E_1 < \sup_{\tilde{\epsilon} \leq u \leq 1} \left[2u^d \left(\eta_{\max} + \frac{d}{2} \frac{1}{u^d} \log(1 - u^2) \right) \right] + \delta$$

for n large enough, where $\tilde{\epsilon} = (\epsilon/\eta_{\max})^{1/d}$. Recalling (4) and choosing δ small enough, we deduce that the condition $\eta_{\max} < \frac{d}{2} (\beta_d^{2nd})^2$ implies that $E_1 \rightarrow 0$. This holds for any order $d \geq 2$.

The E_2 term. The Varadhan lemma may be invoked: but its conclusion, namely $\frac{1}{n} \log E_2 \rightarrow 0$, says nothing on the boundedness of E_2 . We have, however

$$\begin{aligned} E_2 &= \int_0^\infty \mathbb{P}(\exp(2n\eta) \geq t \text{ and } \eta \leq \epsilon) dt \\ &= \int_{-\infty}^0 \mathbb{P}(\eta \geq u \text{ and } \eta \leq \epsilon) 2n \exp(2nu) du + \\ &\quad \int_0^\epsilon \mathbb{P}(\eta \geq u \text{ and } \eta \leq \epsilon) 2n \exp(2nu) du \\ &\leq \mathbb{P}(\eta \leq \epsilon) + \int_0^\epsilon \mathbb{P}(\eta \geq u) 2n \exp(2nu) du. \end{aligned}$$

A weak consequence of the LDP on η is the concentration of η around 0, namely $\mathbb{P}(\eta \leq \epsilon) = 1 - \mathbb{P}(\eta > \epsilon) = 1 - o(1)$. We recall the expanded expression for η : see (2). Notice that $\eta \geq u$ implies that at least one of the r^2 terms of this expansion is at least equal to $\frac{u}{r^2}$. By the union bound, and the fact that $\Re \prod_{k=1}^d \xi_k^{(i,j)} \leq \prod_{k=1}^d |\xi_k^{(i,j)}|$ we deduce that $\mathbb{P}(\eta \geq u) \leq \sum_{i,j=1}^r \mathbb{P}\left(\prod_{k=1}^d |\xi_k^{(i,j)}| \geq \frac{u}{r^2 \lambda_i \lambda_j}\right)$. Invoking again the union bound and noticing that for fixed i, j , $(\xi_k^{(i,j)})_{k=1,\dots,d}$ have the same distribution, we deduce that

$$\mathbb{P}(\eta \geq u) \leq d \sum_{i,j=1}^r \mathbb{P}\left(|\xi_k^{(i,j)}| \geq \left(\frac{u}{r^2 \lambda_i \lambda_j}\right)^{1/d}\right).$$

Now, the density of $\xi_k^{(i,j)}$ is in polar coordinates $\frac{n-1}{\pi} (1-r^2)^{n-2}$ hence, choosing ϵ such that $\epsilon \leq r^2 \max_{i,j} \lambda_i \lambda_j$:

$$\mathbb{P}\left(|\xi_k^{(i,j)}| \geq \left(\frac{u}{r^2 \lambda_i \lambda_j}\right)^{1/d}\right) = \left(1 - \left(\frac{u}{r^2 \lambda_i \lambda_j}\right)^{2/d}\right)^{n-1}.$$

For any $0 \leq x < 1$, $\log(1-x) \leq -x$, hence

$$E_2 \leq d \sum_{i,j} 2n \int_0^\epsilon \exp\left(-(n-1) \left(\frac{u}{r^2 \lambda_i \lambda_j}\right)^{2/d} + 2nu\right) du.$$

When $d > 2$, it is always possible to determine ϵ sufficiently small such that $-(n-1) \left(\frac{u}{r^2 \lambda_i \lambda_j}\right)^{2/d} + 2nu \leq -\frac{n-1}{2} \left(\frac{u}{r^2 \lambda_i \lambda_j}\right)^{2/d}$. This implies that, for such an ϵ , we have

$$E_2 \leq d^2 r^2 n \left(\frac{2}{n-1}\right)^{d/2} \sum_{i,j} \lambda_i \lambda_j \int_0^\infty v^{d/2-1} \exp(-v) dv.$$

The r.h.s. is of course $o(1)$ since $d > 2$.

Remark 7. The bound $\sqrt{\eta_{\max}} < \sqrt{\frac{d}{2}} \beta_d^{2\text{nd}}$ guarantees the non-detectability but it is not tight in general because, in order to study the asymptotics of E_1 , we replaced the true GRF I_η by the lower bound (8). Based on the loose inequality $\log \det(\mathbf{I}_r - \psi_k^* \psi_k) \leq \log \det(1 - \|\psi_k\|_2^2)$, (8) may not be very accurate. It is easy to check that the equality is reached in (8) when all the matrices $(\chi_0^{(k)})_{k=1,\dots,d}$ are rank 1, i.e. if the rank of $\mathbf{X}_0$ is equal to 1. Therefore, the lower bound (8) of I_η is all the better as all the matrices $\chi_0^{(k)}$ are close to being rank 1 matrices. This suggests that, conversely, the bound (8) is likely to be loose when matrices $(\chi_0^{(k)})_{k=1,\dots,d}$ are close to be orthogonal. As an illustration, we would like to consider experimental results. For a given configuration of the spike, we have chosen at random the matrices ψ_k

with $\|\psi_k\| \leq 1$. For each trial, we plot the points of coordinates $x = \eta(\psi_1, \dots, \psi_d)$ and $y = \sum_{k=1}^d \log \det(\mathbf{I}_r - \psi_k^* \psi_k)$ and we obtain a cloud the upper envelope of which is a representation of the true GRF of η ; for comparison, we have plotted the graph of the function defined by the lower bound (8). We have chosen $r = 2$, $d = 3$, and two configurations of the spike: in the first one, all the matrices χ_k have orthogonal columns (top graph of 1), in the second one, the eigenvalues of $\chi_k^* \chi_k$ are the same for $k = 1, 2$ equal to 1.8 and 0.2 (bottom graph of 1).

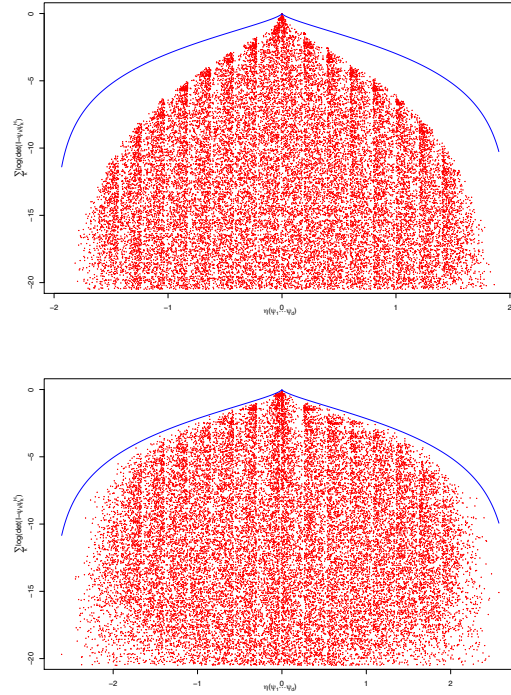


Fig. 1. $-I_\eta$ (seen as the upper envelope of the point constellation) and our upper bound

Remark 8. In the specific case $d = 2$, it is possible to compute in closed-form the exact GRF I_η of η , and to establish the following result: if $\mu_{\max}(\mathbf{X}_0 \mathbf{X}_0^*) < \beta_2^{2\text{nd}} = 1$ (here, $\mu_{\max}$ denotes the largest eigenvalue), then E_1 converges towards 0 and E_2 is bounded: see [5].

6. CONCLUSION

In this paper, we have addressed the detection problem of a rank r high-dimensional tensor $\mathbf{X}_0$. We have generalized the results of [10] to the case where $r > 1$, and established that if the parameter $\eta_{\max}$ defined by (6) is less than parameter $\beta_2^{2\text{nd}}$ introduced in [10], the low rank tensor is undetectable. This condition is based on the lower bound (8) of the GRF I_η which is however not tight in general. It is thus relevant to try to improve this bound in a future work.

7. REFERENCES

- [1] J. Banks, C. Moore, R. Vershynin, N. Verzelen, and J. Xu. Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1137–1141, June 2017.
- [2] Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.
- [3] P. Bianchi, M. Debbah, M. Maïda, and M. Najim. Performance of statistical tests for single source detection using random matrix theory. *IEEE Transactions on Information Theory*, 57(4):2400–2419, 2011.
- [4] Wei-Kuo Chen. Phase transition in the spiked random tensor with rademacher prior. *arXiv:1712.01777v2 [math.PR]*, 2017.
- [5] Antoine Chevreuil and Philippe Loubaton. On the detection of low rank matrices in the high-dimensional regime. <http://arxiv.org/abs/1804.04851>, 2018.
- [6] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Springer-Verlag Berlin Heidelberg, 2009.
- [7] Fabrice Gamboa and Alain Rouault. Operator-valued spectral measures and large deviations. *J. of Stat. Planning and Inference*, 154(3):72–86, Nov. 2014.
- [8] S. Hopkins, J. Shi, and D. Steurer. Tensor principal component analysis via sum-of-squares proofs. In *JMLR: Workshop and Conference Proc.*, volume 40, pages 1–51, 2015.
- [9] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515, June 2017.
- [10] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: from the gaussian hidden clique problem to rank one perturbations of gaussian tensors. *IEEE Trans. Inf. Theor.*, 63(3):1572–1579, March 2017.
- [11] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’14, pages 2897–2905, Cambridge, MA, USA, 2014. MIT Press.
- [12] R.R. Nadakuditi and A. Edelman. Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions on Signal Processing*, 56(7):2625–2637, 2008.
- [13] A. Onatski, M.J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statistics*, 41(3):1204–1231, 2013.
- [14] A. Perry, A.S. Wein, and A. Bandeira. Statistical limits of spiked tensor models. *arXiv:1612.07728v2 [math.PR]*, 12 2016.
- [15] Michel Talagrand. *Mean Field Models for Spin Glasses Book Subtitle. Volume I: Basic Examples*. Springer-Verlag Berlin Heidelberg, 2010.