



**HAL**  
open science

# Large deviation analysis of the CPD detection problem based on random tensor theory

Remy Boyer, Philippe Loubaton

► **To cite this version:**

Remy Boyer, Philippe Loubaton. Large deviation analysis of the CPD detection problem based on random tensor theory. 25th European Signal Processing Conference (EUSIPCO 2017), Aug 2017, Kos Island, Greece. 10.23919/eusipco.2017.8081289. hal-01572144

**HAL Id: hal-01572144**

**<https://centralesupelec.hal.science/hal-01572144v1>**

Submitted on 4 Aug 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.

# LARGE DEVIATION ANALYSIS OF THE CPD DETECTION PROBLEM BASED ON RANDOM TENSOR THEORY

Rémy Boyer

L2S - Department of Signals and Statistics  
Université Paris-Sud, CNRS, CentraleSupélec  
France

remy.boyer@l2s.centralesupelec.fr

Philippe Loubaton

LIGM  
Université de Marne la Vallée  
France

philippe.loubaton@u-pem.fr

## ABSTRACT

The performance in terms of minimal Bayes' error probability for detection of a random tensor is a fundamental understudied difficult problem. In this work, we assume that we observe under the alternative hypothesis a noisy rank- $R$  tensor admitting a  $Q$ -order Canonical Polyadic Decomposition (CPD) with large factors of size  $N_q \times R$ , *i.e.*, for  $1 \leq q \leq Q$ ,  $R, N_q \rightarrow \infty$  with  $R^{1/q}/N_q$  converges to a finite constant. The detection of the random entries of the core tensor is hard to study since an analytic expression of the error probability is not easily tractable. To mitigate this technical difficulty, the Chernoff Upper Bound (CUB) and the error exponent on the error probability are derived and studied for the considered tensor-based detection problem. These two quantities are related to a key quantity for the considered detection problem due to its strong link with the moment generating function of the log-likelihood test. However, the tightest CUB is reached for the value, denoted by  $s^*$ , which minimizes the error exponent. To solve this step, two methodologies are standard in the literature. The first one is based on the use of a costly numerical optimization algorithm. An alternative strategy is to consider the Bhattacharyya Upper Bound (BUB) for  $s^* = 1/2$ . In this last scenario, the costly numerical optimization step is avoided but no guaranty exists on the optimality of the BUB. Based on powerful random matrix theory tools, a simple analytical expression of  $s^*$  is provided with respect to the Signal to Noise Ratio (SNR) and for low rank CPD. Associated to a compact expression of the CUB, an easily tractable expression of the tightest CUB and the error exponent are provided and analyzed. A main conclusion of this work is that the BUB is the tightest bound at low SNRs. At contrary, this property is no longer true for higher SNRs.

## 1. INTRODUCTION

Detection of random parameters in noise is well-known to be a difficult problem. Indeed, the optimal Bayes' decision rule can often only be derived at the price of a costly numerical computation of the log posterior-odds ratio and an exact cal-

culuation of the minimal Bayes' error probability is often intractable [1,2]. This is particularly true in the context of the understudied tensor detection problem, *i.e.*, when the parameters of interest are multidimensional and random. Specifically, the context of this work is about random core tensor detection when the noise-free tensor follows a Canonical Polyadic Decomposition (CPD) [17] with large factors. Following the same methodology presented in [10] for the detection of one-dimensional data, we exploit well-known geometry information divergence [9,11]. The Chernoff upper bound (CUB) is an upper bounds on the minimal Bayes' error probability and the error exponent characterizes the asymptotic exponentially decay of the Bayes' error probability. These two metrics turn out to be useful in many problems of practical importance as for instance, distributed sparse detection [13], sparse support recovery [14], energy detection, MIMO radar processing, network secrecy [15], Angular Resolution Limit in array processing [16], detection performance for informed communication systems, *etc.*. The theory of low rank CPD is a timely and important research topic [3,12]. This class of tensor-based model is useful to extract and analyze relevant information confined into a small dimensional subspace from a massive volume of measurements.

The Random Matrix Theory (RMT) provides a powerful formalism to study the asymptotic performance limits of a large scale system [4,6]. While it exists a plethora of well-known results for linear systems, there is a lack of results for *structured* linear systems encounter with the CPD.

## 2. RANDOM TENSOR DETECTION

### 2.1. CPD and noisy structured linear system

#### 2.1.1. Preliminary definitions

The rank- $R$  CPD of order  $Q$  is defined according to

$$\mathcal{X} = \sum_{r=1}^R s_r \underbrace{\left( \phi_r^{(1)} \circ \dots \circ \phi_r^{(Q)} \right)}_{\mathcal{X}_r} \quad \text{with } \text{rank} \mathcal{X}_r = 1$$

where  $\circ$  is the outer product [12],  $\phi_r^{(q)} \in \mathbb{R}^{N_q \times 1}$  and  $s_r$  is a real scalar. An equivalent formulation using the mode product [12] is

$$\mathcal{X} = \mathcal{S} \times_1 \Phi_1 \times_2 \dots \times_Q \Phi_Q \in \mathbb{R}^{N_1 \times \dots \times N_Q} \quad (1)$$

where  $\times_q$  stands for the  $q$ -mode product,  $\mathcal{S}$  is the  $R \times \dots \times R$  diagonal core tensor with  $[\mathcal{S}]_{r, \dots, r} = s_r$  and  $\Phi_q = [\phi_1^{(q)} \dots \phi_R^{(q)}]$  is the  $q$ -th factor matrix of size  $N_q \times R$ . The  $q$ -mode unfolding matrix for tensor  $\mathcal{X}$  is given by

$$\mathbf{X}_{(q)} = \Phi_q \mathbf{S} (\Phi_Q \circ \dots \circ \Phi_{q+1} \circ \Phi_{q-1} \dots \circ \Phi_1)^T \quad (2)$$

where  $\mathbf{S} = \text{diag}(\mathbf{s})$  with  $\mathbf{s} = [s_1 \dots s_R]^T$  and  $\circ$  stands for the Khatri-Rao product.

### 2.1.2. Vectorized CPD

Assume that the multidimensional measurement tensor follows a noisy  $Q$ -order tensor of size  $N_1 \times \dots \times N_Q$  given by

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} \quad (3)$$

where  $\mathcal{N}$  is the noise tensor where each entry is assumed to be centered *i.i.d.* Gaussian, *i.e.*  $[\mathcal{N}]_{n_1, \dots, n_Q} \sim \mathcal{N}(0, \sigma^2)$  and the noise-free tensor  $\mathcal{X}$  follows a rank- $R$  CPD defined in eq. (1). We assume that each diagonal entry of the core tensor is centered *i.i.d.* Gaussian, *i.e.*  $[\mathcal{S}]_{r, \dots, r} = s_r \sim \mathcal{N}(0, \sigma_s^2)$ . Let  $N = N_1 \dots N_Q$ . The vectorization of eq. (3) is given by

$$\mathbf{y}_N = \text{vec} \mathbf{Y}_{(1)} = \mathbf{x} + \mathbf{n} \quad (4)$$

where  $\mathbf{n} = \text{vec} \mathbf{N}_{(1)}$  and  $\mathbf{x} = \text{vec} \mathbf{X}_{(1)}$ . Using eq. (2) and property  $\text{vec}\{\mathbf{A} \text{diag}\{\mathbf{c}\} \mathbf{B}^T\} = (\mathbf{B} \circ \mathbf{A})\mathbf{c}$ , we obtain

$$\mathbf{x} = \text{vec} \{ \Phi_1 \mathbf{S} (\Phi_Q \circ \dots \circ \Phi_2)^T \} = \Phi \mathbf{s}$$

where  $\Phi = \Phi_Q \circ \dots \circ \Phi_1$  is a  $N \times R$  structured matrix. At this point, it is important to note that the CPD formalism implies that vector  $\mathbf{x}$  in eq. (4) is related to the structured linear system  $\Phi \mathbf{s}$ .

## 2.2. Binary hypothesis test formulation for structured linear system

### 2.2.1. Formulation based on a SNR-type criterion

Let  $\text{SNR} = \sigma_s^2 / \sigma^2$  and  $p_i(\cdot) = p(\cdot | \mathcal{H}_i)$  with  $i \in \{0, 1\}$ . The equi-probable binary hypothesis test for the detection of the random signal,  $\mathbf{s}$ , is

$$\begin{cases} \mathcal{H}_0 : p_0(\mathbf{y}_N; \Phi, \text{SNR} = 0) = \mathcal{N}(\mathbf{0}, \Sigma_0), \\ \mathcal{H}_1 : p_1(\mathbf{y}_N; \Phi, \text{SNR} \neq 0) = \mathcal{N}(\mathbf{0}, \Sigma_1) \end{cases} \quad (5)$$

where  $\Sigma_0 = \sigma^2 \mathbf{I}_N$  and  $\Sigma_1 = \sigma^2 (\text{SNR} \cdot \Phi \Phi^T + \mathbf{I}_N)$ . The data-space for the null hypothesis ( $\mathcal{H}_0$ ) is given by  $\mathcal{X}_0 = \mathcal{X} \setminus \mathcal{X}_1$  where

$$\mathcal{X}_1 = \left\{ \mathbf{y}_N : \Lambda(\mathbf{y}_N) = \log \frac{p_1(\mathbf{y}_N)}{p_0(\mathbf{y}_N)} > \tau' \right\}$$

is the data-space for the alternative hypothesis ( $\mathcal{H}_1$ ). In the above test,  $\Lambda(\mathbf{y}_N)$  is the log likelihood ratio test and  $\tau'$  is the detection threshold given by the following two expressions:

$$\Lambda(\mathbf{y}_N) = \frac{\mathbf{y}_N^T \Phi (\Phi^T \Phi + \text{SNR} \cdot \mathbf{I})^{-1} \Phi^T \mathbf{y}_N}{\sigma^2},$$

$$\tau' = -\log \det (\text{SNR} \cdot \Phi \Phi^T + \mathbf{I}_N)$$

where  $\det(\cdot)$  and  $\log(\cdot)$  stand for the determinant and the natural logarithm, respectively.

### 2.2.2. Geometry of the expected log-likelihood ratio

Consider  $p(\mathbf{y}_N | \hat{\mathcal{H}}) = \mathcal{N}(\mathbf{0}, \Sigma)$  associated to the estimated hypothesis  $\hat{\mathcal{H}}$ . The expected log-likelihood ratio is given by

$$\begin{aligned} \mathbb{E}_{\mathbf{y}_N | \hat{\mathcal{H}}} \Lambda(\mathbf{y}_N) &= \int_{\mathcal{X}} p(\mathbf{y}_N | \hat{\mathcal{H}}) \log \frac{p_1(\mathbf{y}_N)}{p_0(\mathbf{y}_N)} d\mathbf{y}_N \\ &= \mathcal{KL}(\hat{\mathcal{H}} || \mathcal{H}_0) - \mathcal{KL}(\hat{\mathcal{H}} || \mathcal{H}_1) \\ &= \frac{1}{\sigma^2} \text{Tr} \left\{ (\Phi^T \Phi + \text{SNR} \cdot \mathbf{I})^{-1} \Phi^T \Sigma \Phi \right\} \end{aligned}$$

where

$$\mathcal{KL}(\hat{\mathcal{H}} || \mathcal{H}_i) = \int_{\mathcal{X}} p(\mathbf{y}_N | \hat{\mathcal{H}}) \log \frac{p(\mathbf{y}_N | \hat{\mathcal{H}})}{p_i(\mathbf{y}_N)} d\mathbf{y}_N$$

is the Kullback-Liebler Divergence (KLD) [9]. The expected log-likelihood ratio test admits to a simple geometric characterization based on the difference of two KLDs [2]. But, the performance of the detector of interest in terms of the minimal Bayes' error probability, denoted by  $P_e^{(N)}$ , is quite often difficult to determine analytically [1,2].

## 3. CHERNOFF UPPER BOUND (CUB) AND ERROR EXPONENT

Define the minimal Bayes' error probability conditionally to vector  $\mathbf{y}_N$  according to

$$\Pr(\text{Error} | \mathbf{y}_N) = \frac{1}{2} \min\{P_{1,0}, P_{0,1}\}$$

where  $P_{i,i'} = \Pr(\mathcal{H}_i | \mathbf{y}_N \in \mathcal{X}_{i'})$ . The (average) minimal Bayes' error probability defined by  $P_e^{(N)} = \mathbb{E} \Pr(\text{Error} | \mathbf{y}_N)$  is upper bounded according to the CUB [11] such as

$$P_e^{(N)} \leq \frac{1}{2} \cdot \exp[-\mu_N(s)] \quad (6)$$

where the (Chernoff)  $s$ -divergence for  $s \in (0, 1)$  is given by

$$\mu_N(s) = -\log M_{\Lambda(\mathbf{y}_N | \mathcal{H}_1)}(-s) \quad (7)$$

in which  $M_X(t) = \mathbb{E} \exp[t \cdot X]$  is the moment generating function (mgf) of variable  $X$ . The error exponent, denoted by

$\mu(s)$ , is given by the Chernoff information is an asymptotic characterization on the exponentially decay of the minimal Bayes' error probability. The error exponent is derived thanks to the Stein's lemma according to [18]

$$-\lim_{N \rightarrow \infty} \frac{\log P_e^{(N)}}{N} = \lim_{N \rightarrow \infty} \frac{\mu_N(s)}{N} \stackrel{\text{def.}}{=} \mu(s).$$

As parameter  $s \in (0, 1)$  is free, the CUB can be tighter thanks the unique minimizer:

$$s^* = \arg \min_{s \in (0,1)} \mu(s). \quad (8)$$

Finally using eq. (6) and eq. (8), we obtain the Chernoff Upper Bound (CUB). The Bhattacharyya Upper Bound (BUB) is obtained by eq. (6) and by fixing  $s = 1/2$  instead of solving eq. (8). We have the following relation of order:

$$P_e^{(N)} \leq \frac{1}{2} \cdot \exp[-\mu_N(s^*)] \leq \frac{1}{2} \cdot \exp[-\mu_N(1/2)].$$

### 3.1. Error exponent expression based on the RMT

#### 3.1.1. The CUB and the error exponent

In this section, we first recall in the following Lemma the closed-form expression of the CUB for the test of eq. (5). Next, the error exponent,  $\mu(s)$ , is derived in Result 3.2 in the RMT context.

**Lemma 3.1 ([10])** *The log-mgf given by eq. (7) for test of eq. (5) is given by*

$$\begin{aligned} \mu_N(s) &= \frac{1-s}{2} \log \det \left( \text{SNR} \cdot \Phi \Phi^T + \mathbf{I} \right) \\ &\quad - \frac{1}{2} \log \det \left( \text{SNR} \cdot (1-s) \Phi \Phi^T + \mathbf{I} \right). \end{aligned}$$

In the following, we assume that matrices  $(\Phi_q)_{q=1,\dots,Q}$  are random matrices with Gaussian  $\mathcal{N}(0, \frac{1}{N_q})$  entries, an evaluate the behaviour  $\mu_N(s)/N$  when  $(N_q)_{q=1,\dots,Q}$  converge towards  $+\infty$  at the same rate and that  $\frac{R}{N}$  converges towards a non zero limit.

**Result 3.2** *In the asymptotic regime where  $N_1, \dots, N_Q$  converge towards  $+\infty$  at the same rate and where  $R \rightarrow +\infty$  in such a way that  $c_R = \frac{R}{N}$  converges towards a finite constant  $c > 0$ , it holds that*

$$\begin{aligned} \frac{\mu_N(s)}{N} &\xrightarrow{\text{a.s.}} \mu(s) \\ &= \frac{1-s}{2} \Psi_c(\text{SNR}) - \frac{1}{2} \Psi_c((1-s) \cdot \text{SNR}) \quad (9) \end{aligned}$$

with a.s standing for "almost sure convergence" and

$$\begin{aligned} \Psi_c(x) &= \log \left( 1 + \frac{2c}{u(x) + (1-c)} \right) \\ &\quad + c \cdot \log \left( 1 + \frac{2}{u(x) - (1-c)} \right) \\ &\quad - \frac{4c}{x(u(x)^2 - (1-c)^2)} \quad (10) \end{aligned}$$

with  $u(x) = \frac{1}{x} + \sqrt{(\frac{1}{x} + \lambda_c^+)(\frac{1}{x} + \lambda_c^-)}$  where  $\lambda_c^\pm = (1 \pm \sqrt{c})^2$ .

**Proof** See Appendix 6.1.

#### 3.1.2. Approximated analytical expressions for $c \ll 1$

For low rank CPD we have  $R \ll N$  and thus it is realistic to assume  $c \ll 1$ .

**Result 3.3** *In this context, the error exponent can be approximated according to*

$$\mu(s) \stackrel{c \ll 1}{\approx} \frac{c}{2} \left( (1-s) \log(1 + \text{SNR}) - \log(1 + (1-s)\text{SNR}) \right).$$

**Proof** See Appendix 6.2.

As the second-order derivative of  $\mu(s)$  is strictly positive,  $\mu(s)$  is a strictly convex function over interval  $(0, 1)$ . In addition, as a strictly convex function has at most one global minimum, we deduce that the stationary point  $s^*$  is a global minimizer and is given by zeroing the first-order derivative of the error exponent. This optimal value is given by

$$s^* \stackrel{c \ll 1}{\approx} 1 + \frac{1}{\text{SNR}} - \frac{1}{\log(1 + \text{SNR})}. \quad (11)$$

We can identify the two following limit scenarios:

- At low SNR, the error exponent associated with the tightest CUB, denoted by  $\mu(s^*)$ , coincides with the error exponent associated with the BUB. Indeed, the optimal value in eq. (11) admits a second-order approximation for  $c \ll 1$  according to

$$s^* \stackrel{2}{\approx} 1 + \frac{1}{\text{SNR}} \left( 1 - \left( 1 + \frac{\text{SNR}}{2} \right) \right) = \frac{1}{2}.$$

Using Result 3.2 and the above approximation, the best error exponent at low SNR and for  $c \ll 1$  is given by

$$\begin{aligned} \mu \left( \frac{1}{2} \right) &\stackrel{\text{SNR} \ll 1}{\approx} \frac{1}{4} \Psi_{c \ll 1}(\text{SNR}) - \frac{1}{2} \Psi_{c \ll 1} \left( \frac{\text{SNR}}{2} \right) \\ &= \frac{c}{2} \log \frac{\sqrt{1 + \text{SNR}}}{1 + \frac{\text{SNR}}{2}}. \end{aligned}$$

- At contrary for  $\text{SNR} \rightarrow \infty$ , we have  $s^* \rightarrow 1$ . So, the error exponent associated to BUB cannot be considered as optimal in this regime. Using eq. (11) in Result 3.3 and assuming that  $\frac{\log \text{SNR}}{\text{SNR}} \rightarrow 0$ , the optimal error exponent for large SNR can be approximated according to

$$\mu(s^*) \stackrel{\text{SNR} \ll 1}{\approx} \frac{c}{2} (1 - \log \text{SNR} + \log \log(1 + \text{SNR})).$$

## 4. NUMERICAL SIMULATIONS

In this simulation part, we consider a cubic tensor of order  $Q = 3$  with  $N_1 = N_2 = N_3 = 100$ . The factors  $\Phi_1, \Phi_2$  and  $\Phi_3$  are generated as a single *i.i.d.* Gaussian realization of rank  $R = 20$ . We can check  $c = 2e - 5 \ll 1$ . In Fig. 1, it is drawn parameter  $s^*$  with respect to the SNR in dB. The parameter  $s^*$  is obtained thanks to three different methods. The first one is based on the brut force/exhaustive computation of the CUB thanks to a discretization of the  $s$  parameter over a fine grid. This approach has two drawbacks. First,  $s^*$  cannot be estimated with an accuracy depending to the grid precision. The second problem is that this exhaustive procedure has a very high computational cost especially in our asymptotic context. The second approach is based on the numerical optimization of the closed-form of  $\mu(s)$  given in Result 3.3. In this scenario, the drawback in terms of the computational cost is mitigated but the precision drawback due to the grid design remains. Finally, the last approach is based on the analytic expression given in eq. (11) under the hypothesis that  $c \ll 1$ . This last strategy has a negligible computational cost and does not suffer from the grid precision limitation. We can check that the three methods coincide with a high accuracy. We can also verify the limit values of  $s^*$  given in section 3.1.2 in the small and large SNR regimes. In Fig. 2, the same three scenarios are considered. Here again, we can observe the good agreement of the three approaches.

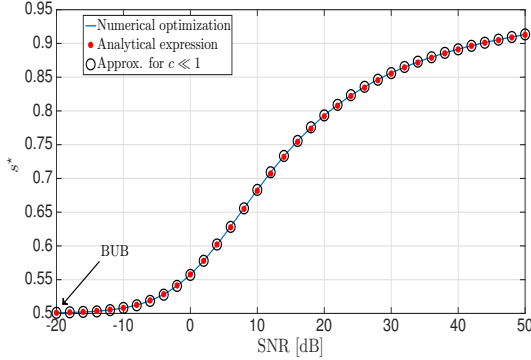


Fig. 1. Optimal  $s$ -parameter vs. SNR in dB

## 5. CONCLUSION

The derivation and the analyze of the asymptotic performance in terms of minimal Bayes' error probability for the detection of a random parameters is addressed in this work. More precisely, we assume that under te alternative hypothesis, we observe a noisy  $Q$ -order tensor following a rank- $R$  CPD with large factors and an unknown random core tensor of interest. The term "large" means that the number of available measurements,  $N$ , and the number of desired random parameters,  $R$ ,

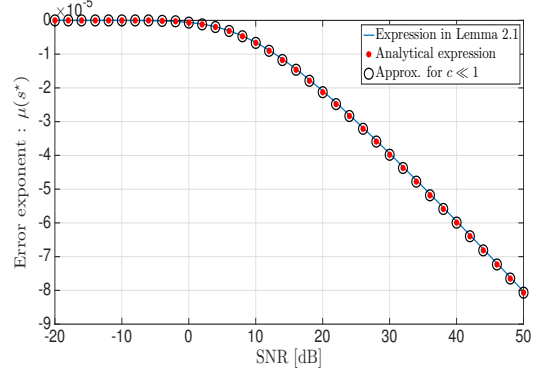


Fig. 2. Error exponent vs. SNR in dB

grow jointly to infinity with an asymptotically constant ratio. The CUB and the error exponent are proposed in closed-form. In addition, it is provided analytical expressions of the optimal parameter  $s$  for which the CUB is a tight upper bound on the Bayes' error probability.

## 6. APPENDIX

### 6.1. Proof of Result 3.2

Large random matrix theory allows to evaluate the asymptotic behavior of  $\frac{\mu_N(s)}{N}$  when  $N_q \rightarrow +\infty$  for each  $q = 1, \dots, Q$ ,  $R \rightarrow +\infty$  in such a way that  $\frac{R^{1/q}}{N_q}$  converge towards a non zero constant for each  $q = 1, \dots, Q$ . In other words,  $N_1, \dots, N_Q$  converge towards  $+\infty$  at the same rate (i.e.  $\frac{N_q}{N_p}$  converge towards a non zero constant for each  $(p, q)$ ), and  $c_R = \frac{R}{N}$  converges towards a constant  $c > 0$ . In this context, the empirical eigenvalue distribution of matrix  $\Phi\Phi^T$  converges towards a relevant Marcenko-Pastur distribution. More precisely, we define the Marcenko-Pastur distribution  $\mu_c(d\lambda)$  as the probability distribution given by

$$\mu_c(d\lambda) = \delta(\lambda) [1-c]_+ + \frac{\sqrt{(\lambda - \lambda_c^-)(\lambda_c^+ - \lambda)}}{2\pi\lambda} \mathbb{1}_{[\lambda_c^-, \lambda_c^+]}(\lambda) d\lambda$$

where  $\lambda_c^- = (1 - \sqrt{c})^2$  and  $\lambda_c^+ = (1 + \sqrt{c})^2$ . The Stieltjes transform of  $\mu_c$  defined as  $t_c(z) = \int_{\mathbb{R}^+} \frac{\mu_c(d\lambda)}{\lambda - z}$  is known to satisfy the equation

$$t_c(z) = \left[ -z + \frac{c}{1 + t_c(z)} \right]^{-1}.$$

When  $z \in \mathbb{R}^{-*}$ , i.e.  $z = -\rho$ , with  $\rho > 0$ , it is well known that  $t_c(\rho)$  is given by

$$t_c(-\rho) = \frac{2}{\rho - (1 - c) + \sqrt{(\rho + \lambda_c^-)(\rho + \lambda_c^+)}} \quad (12)$$

It was established for the first time in [6] that if  $\mathbf{X}$  represents a  $M \times P$  random matrix with zero mean and  $\frac{1}{M}$  variance i.i.d. entries, and if  $(\lambda_m)_{m=1,\dots,M}$  represent the eigenvalues of  $\mathbf{X}\mathbf{X}^T$  arranged in decreasing order, then the so-called empirical eigenvalue distribution of  $\mathbf{X}\mathbf{X}^T$  defined as  $\frac{1}{M} \sum_{m=1}^M \delta(\lambda - \lambda_m)$  converges weakly almost surely towards  $\mu_c$  in the asymptotic regime where  $M \rightarrow +\infty$ ,  $P \rightarrow +\infty$ ,  $\frac{P}{M} \rightarrow c$ . In particular, for each continuous function  $f(\lambda)$ , it holds that

$$\frac{1}{M} \sum_{m=1}^M f(\lambda_m) \xrightarrow{\text{a.s.}} \int_{\mathbb{R}^+} f(\lambda) \mu_c(d\lambda). \quad (13)$$

In practice, this result means that if  $M$  and  $K$  are large enough, then the histogram of the eigenvalues of each realization of  $\mathbf{X}\mathbf{X}^T$  tends to accumulate around the graph of the probability density of  $\mu_c$ .

The columns  $(\phi_r)_{r=1,\dots,R}$  of  $\Phi$  are vectors  $(\phi_r^{(Q)} \otimes \dots \otimes \phi_r^{(1)})_{r=1,\dots,R}$ . These vectors are mutually independent, identically distributed, and satisfy  $\mathbb{E}(\phi_r \phi_r^T) = \frac{I_N}{N}$ . However, the elements of  $\Phi$  are not mutually independent because the components of each column  $\phi_r$  are not independent. In the asymptotic regime considered in this paper, the results of [8] (see also [5]) allow to establish that the empirical eigenvalue distribution of  $\Phi\Phi^T$  still converges almost surely towards  $\mu_c$ , where we recall that  $\frac{R}{N} \rightarrow c$ . Using (13) for  $f(\lambda) = \log(1 + \lambda/\rho)$  as well as a well-known formula that allows to express  $\int_{\mathbb{R}^+} \log(1 + \lambda/\rho) \mu_c(d\lambda)$  in terms of  $t_c(-\rho)$  given by (12) (see e.g. [7]), we obtain the following result.

## 6.2. Proof of Result 3.3

We have  $u(x) \approx \frac{1}{x} + \sqrt{(\frac{1}{x} + 1)^2} = \frac{2}{x} + 1$  and  $u(x) + (1-c) \approx 2(\frac{1}{x} + 1)$ ,  $u(x) - (1-c) \approx \frac{2}{x}$ ,  $u(x)^2 - (1-c)^2 \approx \frac{4}{x}(\frac{1}{x} + 1)$ . Using the above first-order approximations, eq. (10) is

$$\Psi_{c \ll 1}(x) \approx c \cdot \frac{x}{1+x} + c \log(1+x) - c \frac{x}{1+x} = c \log(1+x).$$

Using the above approximation and eq. (9), we obtain Result 3.3.

## REFERENCES

- [1] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory*, PTR Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [2] Y. Cheng, X. Hua, H. Wang, Y. Qin and X. Li, "The Geometry of Signal Detection with Applications to Radar Signal Processing," *Entropy*, 18(11), 381.
- [3] P. Comon, "Tensors: a brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44-53, 2014.
- [4] Z. Bai, J.W. Silverstein, "Spectral analysis of large dimensional random matrices", *Springer Series in Statistics*, 2nd ed., 2010.
- [5] A. Ambainis, A.W. Harrow and M.B. Hastings, "Random matrix theory: extending random matrix theory to mixtures of random product states", *Commun. Math. Phys.*, vol. 310, no. 1, pp. 25-74 (2012)
- [6] V. A. Marchenko, and L.A. Pastur, "Distribution of eigenvalues for some sets of random matrices", *Mat. Sb. (N.S.)*, 72(114):4, 507-536, 1967.
- [7] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*, Hanover, MA, USA: Now Publishers Inc., Jun. 2004, vol. 1, no. 1.
- [8] A. Pajor and L.A. Pastur, "On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution, *Studia Math*, 195 (2009), pp: 11-29.
- [9] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012
- [10] R. Boyer and F. Nielsen, "Information Geometry Metric for Random Signal Detection in Large Random Sensing Systems," *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, 2017
- [11] F. Nielsen, "An information-geometric characterization of Chernoff information," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 269-272, 2013.
- [12] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145-163, 2015.
- [13] S. P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1446-1460, 2015.
- [14] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE transactions on information theory*, vol. 56, no. 3, pp. 1383-1399, 2010.
- [15] R. Boyer and C. Delpha, "Relative-entropy based beamforming for secret key transmission," *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2012
- [16] N. D. Tran, R. Boyer, S. Marcos, and P. Larzabal, "Angular Resolution Limit for array processing: Estimation and information theory approaches," *European Signal Processing Conference (EUSIPCO)*, Aug 2012
- [17] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory? multi-modal factor analysis," *UCLA Working Papers Phon.*, vol. 16, pp. 1-84, 1970.
- [18] S. Sinanovic and D. H. Johnson, "Toward a theory of information processing," *Signal Processing*, vol. 87, no. 6, pp. 1326- 1344, 2007.