Article

Computational Information Geometry For Binary Classification of High-Dimensional Random Tensors †

Gia-Thuy Pham [1], Rémy Boyer [1] and Frank Nielsen [2,3,*]

[1] University of Paris-Sud, L2S, Department of Signals and Statistics, France; giathuy.pham@l2s.centralesupelec.fr (G.-T.P.); remy.boyer@l2s.centralesupelec.fr (R.B.)
[2] École Polytechnique, LIX, France
[3] Sony Computer Science Laboratories, Japan
* Correspondence: Frank.Nielsen@acm.org; Tel.: +x-xxx-xxx-xxxx
† The results presented in this work have been partially published in [16] and [15]

Abstract: The performance in terms of minimal Bayes' error probability for detection of a high-dimensional random tensor is a fundamental under-studied difficult problem. In this work, we consider two Signal to Noise Ratio (SNR)-based detection problems of interest. Under the alternative hypothesis, *i.e.*, for a non-zero SNR, the observed signals are either 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 $1 \leq q \leq Q$, where $R, N_q \to \infty$ with $R^{1/q}/N_q$ converge towards a finite constant or a noisy tensor admitting TucKer Decomposition (TKD) of multilinear $(M_1, \ldots, M_Q)$-rank with large factors of size $N_q \times M_q$, *i.e*, for $1 \leq q \leq Q$, where $N_q, M_q \to \infty$ with M_q/N_q converge towards a *finite* constant. The detection of the random entries (coefficients) of the core tensor in the CPD/TKD is hard to study since the exact derivation of the error probability is mathematically intractable. To circumvent this technical difficulty, the Chernoff Upper Bound (CUB) for larger SNR and the Fisher information at low SNR are derived and studied, based on information geometry theory. The tightest CUB is reached for the value minimizing the error exponent, denoted by $s^\star$. In general, due to the asymmetry of the s-divergence, the Bhattacharyya Upper Bound (BUB) (that is, the Chernoff Information calculated at $s^\star = 1/2$) can not solve this problem effectively. As a consequence, we rely on a costly numerical optimization strategy to find $s^\star$. However, thanks to powerful random matrix theory tools, a simple analytical expression of $s^\star$ is provided with respect to the Signal to Noise Ratio (SNR) in the two schemes considered. A main conclusion of this work is that the BUB is the tightest bound at low SNRs. This property is, however, no longer true for higher SNRs.

Keywords: Optimal Bayesian detection, information geometry, minimal error probability, Chernoff/Bhattacharyya upper bound, large random tensor, Fisher information, large random sensing matrix

1. Introduction

1.1. State-of-the-art and problem statement

Evaluating the performance limit for the "Gaussian information plus noise" detection problem is a challenging research topic, see for instance [6,8,9,34,36,39,46]. Given a binary hypothesis problem, the Bayes' decision rule is based on the principle of the largest posterior probability. Specifically, the Bayesian detector chooses the alternative hypothesis $\mathcal{H}_1$ if $\Pr(\mathcal{H}_1|\mathbf{y}) > \Pr(\mathcal{H}_0|\mathbf{y})$ for a given N-dimensional measurement vector $\mathbf{y}$ or the null hypothesis $\mathcal{H}_0$, otherwise. Consequently, the optimal decision rule can often only be derived at the price of a costly numerical computation of the log posterior-odds ratio [34] since an exact calculation of the minimal Bayes' error probability $P_e^{(N)}$ is often intractable [17,34]. To circumvent this problem, it is standard to exploit well-known bounds on $P_e^{(N)}$ based on information theory [2,24,32,42,47]. In particular, the *Chernoff information* [19,40]

is asymptotically (in N) relied to the exponential rate of $P_e^{(N)}$. The Chernoff information turns out to be useful in many problems of practical importance as for instance, distributed sparse detection [18], sparse support recovery [48], energy detection [35], MIMO radar processing [31,45], network secrecy [13], Angular Resolution Limit in array processing [27], detection performance for informed communication systems [33], just to name a few. In addition, the Chernoff information bound can be tight for a minimal s-divergence over parameter $s \in (0,1)$. Generally, this step requires to solve numerically an optimization problem [41] and often leads to a complicated and uninformative expression of the optimal value of s. To circumvent this difficulty, a simplified case of $s = 1/2$ is often used corresponding to the well-known Bhattacharyya divergence [47] at the price of a less accurate prediction of $P_e^{(N)}$. In information geometry, parameter s is often called α, and the s-divergence is the so-called Chernoff α-divergence [41].

The theory of tensor decomposition is a timely and important research topic [20,23]. Tensors are useful to extract relevant information confined into a small dimensional subspace from a massive and multidimentional volume of measurements. In the standard literature, two main families of tensor decomposition are prominent. Namely the Canonical Polyadic Decomposition (CPD) [23] and the Tucker decomposition (TKD)/HOSVD (High-Order SVD) [25,49]. These approaches are two possible multilinear generalization of the Singular Value Decomposition (SVD). A natural generalization to tensors of the usual concept of rank for matrices is called the CPD. The tensorial/canonical rank of a P-order tensor is equal to the minimal positive integer, say R, of unit rank tensors that must be summed up for perfect recovery. A unit rank tensor is the outer product of P vectors. In addition, the CPD has remarkable uniqueness properties [23] and involves only a reduced number of free parameters due to the constraint of minimality on R. Unfortunately, unlike to the matrix case, the set of tensors with fixed (tensorial) rank is not close [21,26]. This singularity implies that the problem of the computation of the CPD is mathematically ill-posed. The consequence is that its numerical computation remains non trivial and is usually done using suboptimal iterative algorithms [22]. Note that this problem can sometimes be avoided by exploiting some natural hidden structures in the physical model [30]. The TKD [49] and the HOSVD [25] are two popular decompositions being an alternative to the CPD. In this case, the notion of tensorial rank is no longer relevant and a new rank definition is used. Specifically, it is standard to use the *multilinear rank* defined as the set of positive integers $\{R_1, \ldots, R_P\}$ where each integer, R_p, is the usual rank of the p-th mode. Its practical construction is non-iterative and optimal in the sense of the Eckart-Young theorem at each mode level. This approach is interesting because it can be computed in real-time [4] or adaptively [12]. Unfortunately, it is shown that the low (multilinear) rank tensor based on this procedure is generally suboptimal [25]. In other words, there does not exist a generalization of the Eckart-Young theorem for tensors of order strictly greater than two!

The detection performance of a multilinear tensor following the CPD and TKD can be derived and studied. It is important to note that the detection theory for tensors is a very under studied research topic. To the best of our knowledge, only the publication [10] tackles this problem in the context of RADAR multidimensional data detection. A major difference with this publication is that their analysis is based on the performance of a low rank detection after matched filtering.
More specifically, we consider two cases where the observations are either (1) a noisy rank-R tensor admitting a Q-order CPD with large factors of size $N_q \times R$, *i.e*, for $1 \leq q \leq Q$, $R, N_q \to \infty$ with $R^{1/q}/N_q$ converging towards a finite constant, or (2) a noisy tensor admitting a TKD of multilinear $(M_1, \ldots, M_Q)$-rank with large factors of size $N_q \times M_q$, *i.e.*, for $1 \leq q \leq Q$, where $N_q, M_q \to \infty$ with M_q/N_q converging towards a finite constant. For zero-mean independent Gaussian core and noise tensors a key discriminative parameter is the Signal to Noise Ratio defined by $\text{SNR} = \sigma_{\mathbf{s}}^2/\sigma^2$ where $\sigma_{\mathbf{s}}^2$ and σ^2 are the variances of the vectorized core and noise tensors, respectively. So, the binary hypothesis test of interest can be described in the following way:

Under the null hypothesis $\mathcal{H}_0$, $\text{SNR} = 0$, meaning that only the noise is present. Conversely, the alternative hypothesis $\mathcal{H}_1$ is based on $\text{SNR} \neq 0$, meaning that there exists a multilinear signal of interest. First note that there exists a lack of contribution dealing with detection performance

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

for tensors. The detection of the random entries of the core tensor is hard to study since the exact derivation of the error probability is intractable. To circumvent this technical difficulty, based on computational information geometry theory, we consider the Chernoff Upper Bound (CUB), and the Fisher information in the context of massive measurement vectors. The tightest CUB is reached for the value, denoted by $s^\star$, which minimizes the error exponent. In general, due to the asymmetry of the s-divergence, the Bhattacharyya Upper Bound (BUB) — Chernoff Information calculated at $s^\star = 1/2$— cannot solve this problem effectively. As a consequence, we rely on a costly numerical optimization strategy to find $s^\star$. However, thanks to powerful Random Matrix Theory (RMT) tools, a simple analytical expression of $s^\star$ is provided with respect to the Signal to Noise Ratio (SNR). For low SNR, analytical expressions of the Fisher information are given. Note that the analysis of the Fisher information in the context of the RMT has been only studied in recent contributions [11,14,43] for parameter estimation. For larger SNR, analytic and simple expression of the CUB for the CPD and the TKD are provided.

We note that Random Matrix Theory (RMT) has fascinated both mathematicians and physicists since they were first introduced in mathematical statistics by Wishart in 1928 [54]. After a slow start, the subject gained prominence when Wigner [52] introduced the concept of statistical distribution of nuclear energy levels in 1950. However, it took until 1955 before Wigner [53] introduced ensembles of random matrices. Since then, many important results in RMT were developed and analyzed, see for instance [5,29,38,51] and the references therein. In the last two decades, researches on RMT has been constantly published.

1.2. Paper organisation

The organization of the paper is as follows: In the second section, we introduce some definitions, tensor models, and the Marchenko-Pastur distribution from random matrix theory. The third section is devoted to present Chernoff Information for binary hypothesis test. The fourth section gives the main results on Fisher Information and the Chernoff bound. The numerical simulation results are given in the fifth section. We conclude our work by giving some perspectives in the Section 6. Finally, several proofs of the paper can be found in the appendix. We also give the list of theorem, result, lemma, remark and definitions.

List of Theorems

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

2. Algebra of tensors and Random Matrix Theory (RMT)

In this section, we introduce some useful definitions from tensor algebra and from the spectral theory of large random matrices.

2.1. Multilinear functions

2.1.1. Preliminary definitions

Definition 1. *The Kronecker product of matrices* $\mathbf{X}$ *and* $\mathbf{Y}$ *of size* $I \times J$ *and* $K \times N$*, respectively is given by*

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} [\mathbf{X}]_{11}\mathbf{Y} & \dots & [\mathbf{X}]_{1J}\mathbf{Y} \\ \vdots & & \vdots \\ [\mathbf{X}]_{I1}\mathbf{Y} & \dots & [\mathbf{X}]_{IJ}\mathbf{Y} \end{bmatrix} \in \mathbb{R}^{(IK)\times(JN)}.$$

We have $\text{rank}\{\mathbf{X} \otimes \mathbf{Y}\} = \text{rank}\{\mathbf{X}\} \cdot \text{rank}\{\mathbf{Y}\}$.

Definition 2. *The vectorization* $\text{vec}(\boldsymbol{\mathcal{X}})$ *of a tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M_1 \times \dots \times M_Q}$ *is a vector* $\mathbf{x} \in \mathbb{R}^{M_1 M_2 \dots M_Q}$ *defined as*

$$\mathbf{x}_h = [\boldsymbol{\mathcal{X}}]_{m_1,\dots,m_Q}$$

where $h = m_1 + \sum_{k=2}^{Q}(m_k - 1)M_1 M_2 \dots M_{k-1}$.

Definition 3. *The q-mode product denoted by* $\times_q$ *between a tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M_1 \times \dots \times M_Q}$ *and a matrix* $\mathbf{U} \in \mathbb{R}^{K \times M_q}$ *is denoted by* $\boldsymbol{\mathcal{X}} \times_q \mathbf{U} \in \mathbb{R}^{M_1 \times \dots \times M_{q-1} \times K \times M_{q+1} \times \dots \times M_Q}$ *with*

$$[\boldsymbol{\mathcal{X}} \times_q \mathbf{U}]_{m_1,\dots,m_{q-1},k,m_{q+1},\dots,m_Q} = \sum_{m_q=1}^{M_q} [\boldsymbol{\mathcal{X}}]_{m_1,\dots,m_Q}[\mathbf{U}]_{k,m_q}$$

where $1 \le k \le K$.

Definition 4. *The q-mode unfolding matrix of size* $M_q \times \left(\prod_{k=1,k\neq q}^{Q} M_k\right)$ *denoted by* $\mathbf{X}_{(q)} = \text{unfold}_q(\boldsymbol{\mathcal{X}})$ *of a tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M_1 \times \dots \times M_q}$ *is defined according to*

$$[\mathbf{X}_{(q)}]_{M_q,h} = [\boldsymbol{\mathcal{X}}]_{m_1,\dots,m_Q}$$

where $h = 1 + \sum_{k=1,k\neq q}^{Q}(m_k - 1)\prod_{v=1,v\neq q}^{k-1} M_v$.

2.1.2. Canonical Polyadic Decomposition (CPD)

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

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

where $\circ$ is the outer product [20], $\boldsymbol{\phi}_r^{(q)} \in \mathbb{R}^{N_q \times 1}$ and s_r is a real scalar.
An equivalent formulation using the q-mode product defined in Def. 3 is

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{\Phi}^{(1)} \times_2 \dots \times_Q \boldsymbol{\Phi}^{(Q)}$$

where $\boldsymbol{\mathcal{S}}$ is the $R \times \dots \times R$ diagonal core tensor with $[\boldsymbol{\mathcal{S}}]_{r,\dots,r} = s_r$ and $\boldsymbol{\Phi}^{(q)} = [\boldsymbol{\phi}_1^{(q)} \dots \boldsymbol{\phi}_R^{(q)}]$ is the q-th factor matrix of size $N_q \times R$.

The q-mode unfolding matrix for tensor $\boldsymbol{\mathcal{X}}$ is given by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

$$\boldsymbol{X}_{(q)} = \boldsymbol{\Phi}^{(q)} \boldsymbol{S} \left(\boldsymbol{\Phi}^{(Q)} \odot ... \odot \boldsymbol{\Phi}^{(q+1)} \odot \boldsymbol{\Phi}^{(q-1)} \odot ... \odot \boldsymbol{\Phi}^{(1)} \right)^T$$

where $\boldsymbol{S} = \text{diag}(\boldsymbol{s})$ with $\boldsymbol{s} = [s_1, ..., s_R]^T$ and $\odot$ stands for the Khatri-Rao product [20].

2.1.3. Tucker Decomposition (TKD)

The Tucker tensor model of order Q is defined according to

$$\boldsymbol{\mathcal{X}} = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} ... \sum_{m_Q=1}^{M_Q} s_{m_1 m_2 ... m_Q} \left(\boldsymbol{\phi}_{m_1}^{(1)} \circ \boldsymbol{\phi}_{m_2}^{(2)} \circ \cdots \circ \boldsymbol{\phi}_{m_Q}^{(Q)} \right)$$

where $\boldsymbol{\phi}_{m_q}^{(q)} \in \mathbb{R}^{N_q \times 1}$, $q = 1, ..., Q$ and $s_{m_1 m_2 ... m_Q}$ is a real scalar.
The q-mode product of $\boldsymbol{\mathcal{X}}$ is similar to CPD case, however the q-mode unfolding matrix for tensor $\boldsymbol{\mathcal{X}}$ is slightly different

$$\mathbf{X}_{(q)} = \boldsymbol{\Phi}^{(q)} \mathbf{S}_{(q)} \left(\boldsymbol{\Phi}^{(Q)} \otimes \ldots \otimes \boldsymbol{\Phi}^{(q+1)} \otimes \boldsymbol{\Phi}^{(q-1)} \ldots \otimes \boldsymbol{\Phi}^{(1)} \right)^T$$

where $\mathbf{S}_{(q)} \in \mathbb{R}^{N_q \times N_1 N_2 ... N_{q-1} N_{q+1} ... N_Q}$ the q-mode unfolding matrix of tensor $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\Phi}^{(q)} = [\boldsymbol{\phi}_1^{(q)} ... \boldsymbol{\phi}_{M_q}^{(q)}] \in \mathbb{R}^{N_q \times M_q}$ and $\otimes$ stands for Kronecker product.

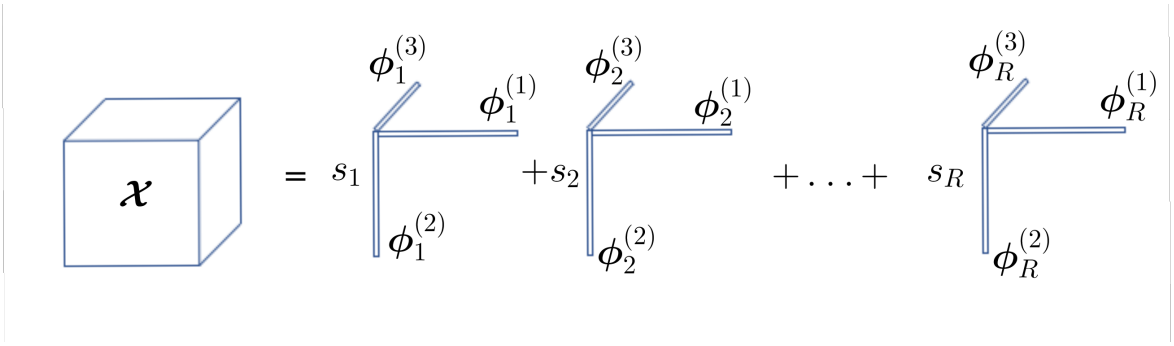

Figure 1. Canonical Polyadic Decomposition (CPD)

At this point, it is important to note that the CPD and TKD formalism implies that vector $\mathbf{x}$ in (11) is related either to the structured linear system $\boldsymbol{\Phi}^{\odot}$ or $\boldsymbol{\Phi}^{\otimes}$.

2.2. The Marchenko-Pastur distribution

The Marchenko-Pastur distribution was introduced half a century ago [38] in 1967, and plays a key role in a number of high-dimensional signal processing problems. To help the reader, in this section, we introduce some fundamental results concerning large empirical covariance matrices. Let $(\boldsymbol{v}_n)_{n=1,...,N}$ a sequence of i.i.d zero mean Gaussian random M-dimensional vectors for which $\mathbb{E}(\boldsymbol{v}_n \boldsymbol{v}_n^T) = \sigma^2 \boldsymbol{I}_M$. We consider the empirical covariance matrix

$$\frac{1}{N} \sum_{n=1}^{N} \boldsymbol{v}_n \boldsymbol{v}_n^T$$

which can be also written as

$$\frac{1}{N} \sum_{n=1}^{N} \boldsymbol{v}_n \boldsymbol{v}_n^T = \boldsymbol{W}_N \boldsymbol{W}_N^T$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

where matrix $\boldsymbol{W}_N$ is defined by $\boldsymbol{W}_N = \frac{1}{\sqrt{N}}[\boldsymbol{v}_1, ..., \boldsymbol{v}_N]$. $\boldsymbol{W}_N$ is thus a Gaussian matrix with independent identically distributed $\mathcal{N}(0, \frac{\sigma^2}{N})$ entries. When $N \to +\infty$ while M remains fixed, matrix $\boldsymbol{W}_N \boldsymbol{W}_N^T$ converges towards $\sigma^2 \boldsymbol{I}_M$ in the spectral norm sense. In the high dimensional asymptotic regime defined by

$$M \to +\infty, \quad N \to +\infty, \quad c_N = \frac{M}{N} \to c > 0$$

it is well understood that $\left\| \boldsymbol{W}_N \boldsymbol{W}_N^T - \sigma^2 \boldsymbol{I}_M \right\|$ does not converge towards 0. In particular, the empirical distribution $\hat{\nu}_N = \frac{1}{M} \sum_{m=1}^{M} \delta_{\hat{\lambda}_{m,N}}$ of the eigenvalues $\hat{\lambda}_{1,N} \geq ... \geq \hat{\lambda}_{M,N}$ of $\boldsymbol{W}_N \boldsymbol{W}_N^T$ does not converge towards the Dirac measure at point $\lambda = \sigma^2$. More precisely, we denote by ν_{c,σ^2} the Marchenko-Pastur distribution of parameters (c, σ^2) defined as the probability measure

$$\nu_{c,\sigma^2}(d\lambda) = \delta_0[1 - \frac{1}{c}]_+ + \frac{\sqrt{(\lambda - \lambda^-)(\lambda^+ - \lambda)}}{2\sigma^2 c \pi \lambda} \mathbb{1}_{[\lambda^-, \lambda^+]}(\lambda) d\lambda \tag{1}$$

with $\lambda^- = \sigma^2(1 - \sqrt{c})^2$ and $\lambda^+ = \sigma^2(1 + \sqrt{c})^2$. Then, the following result holds.

Theorem 5. *([38]) The empirical eigenvalue value distribution $\hat{\mu}_N$ converges weakly almost surely towards μ_{d,σ^2} when both M and N converge towards $+\infty$ in such a way that $c_N = \frac{M}{N}$ converges towards $c > 0$. Moreover, it holds that*

$$\hat{\lambda}_{1,N} \to \sigma^2(1 + \sqrt{c})^2 \ a.s. \tag{2}$$

$$\hat{\lambda}_{\min(M,N)} \to \sigma^2(1 - \sqrt{c})^2 \ a.s. \tag{3}$$

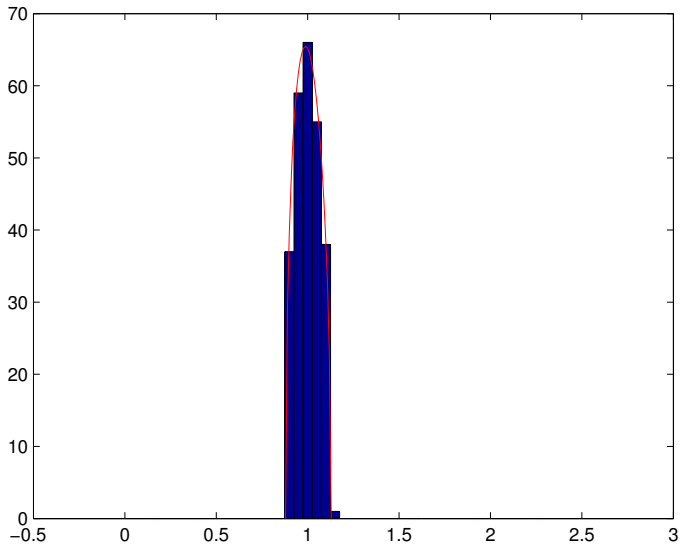

Figure 2. Histogram of the eigenvalues of $\frac{\mathbf{W}_N \mathbf{W}_N^T}{N}$ (with $M = 256, c_N = \frac{M}{N} = \frac{1}{256}, \sigma^2 = 1$)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

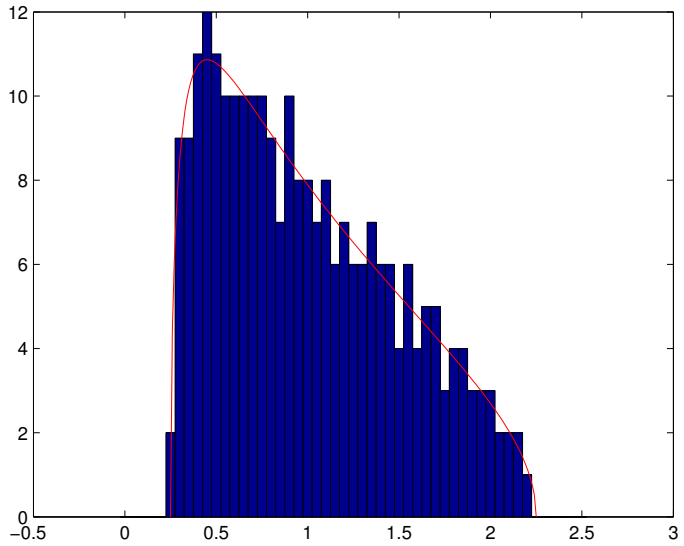

Figure 3. Histogram of the eigenvalues of $\frac{\mathbf{W}_N\mathbf{W}_N^T}{N}$ (with $M = 256, c_N = \frac{M}{N} = \frac{1}{4}, \sigma^2 = 1$)

We also observe that Theorem 5 remains valid if $\mathbf{W}_N$ is not necessarily a Gaussian matrix whose i.i.d. elements have a finite fourth order moment (see e.g. [5]). Theorem 5 means that when ratio $\frac{M}{N}$ is not small enough, the eigenvalues of the empirical spatial covariance matrix of a temporally and spatially white noise tend to spread out around the variance of the noise, and that almost surely, for N large enough, all the eigenvalues are located in a neighbourhood of interval $[\lambda^-, \lambda^+]$.

3. Classification in a Computational Information Geometry (CIG) framework

3.1. Formulation based on a SNR*-type criterion*

Let $\text{SNR} = \sigma_{\mathbf{s}}^2/\sigma^2$ 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

$$\begin{cases} \mathcal{H}_0 : p_0(\mathbf{y}_N; \mathbf{\Phi}, \text{SNR} = 0) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0), \\ \mathcal{H}_1 : p_1(\mathbf{y}_N; \mathbf{\Phi}, \text{SNR} \neq 0) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1) \end{cases} \tag{4}$$

where $\mathbf{\Sigma}_0 = \sigma^2\mathbf{I}_N$ and $\mathbf{\Sigma}_1 = \sigma^2\left(\text{SNR} \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}_N\right)$. 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 τ' is the detection threshold given by the following two expressions:

$$\Lambda(\mathbf{y}_N) = \frac{\mathbf{y}_N^T\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{\Phi} + \text{SNR} \cdot \mathbf{I}\right)^{-1}\mathbf{\Phi}^T\mathbf{y}_N}{\sigma^2},$$
$$\tau' = -\log\det\left(\text{SNR} \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}_N\right)$$

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

3.2. Geometry of the expected log-likelihood ratio

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

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

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) [24]. The expected log-likelihood ratio test admits to a simple geometric characterization based on the difference of two KLDs [17]. 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 [17,34] in closed-form.

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

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

where $P_{i,i'} = \Pr(\mathcal{H}_i|\mathbf{y}_N \in \mathcal{X}_{i'})$.

3.3. CUB

The (average) minimal Bayes' error probability defined by $P_e^{(N)} = \mathbb{E}\Pr(\mathrm{Error}|\mathbf{y}_N)$ is upper bounded according to the CUB [41] such as

$$P_e^{(N)} \leq \frac{1}{2}\cdot\exp[-\mu_N(s)] \tag{5}$$

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

$$\mu_N(s) = -\log M_{\Lambda(\mathbf{y}_N|\mathcal{H}_1)}(-s) \tag{6}$$

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 which 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 [47]

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

As parameter $s \in (0,1)$ is free, the CUB can be tightened by minimizing this parameter:

$$s^\star = \arg\min_{s\in(0,1)} \mu(s). \tag{7}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

Finally using eq. (5) and eq. (7), we obtain the Chernoff Upper Bound (CUB). The Bhattacharyya Upper Bound (BUB) is obtained by eq. (5) and by fixing $s = 1/2$ instead of solving eq. (7). Therefore we have the following relation of order:

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

Lemma 6. *The log-moment generating function given by eq. (6) for test of eq. (4) is given by*

$$\begin{aligned} \mu_N(s) = & \frac{1-s}{2} \log\det\left(\mathrm{SNR} \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \\ & - \frac{1}{2} \log\det\left(\mathrm{SNR} \cdot (1-s)\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right). \end{aligned} \tag{8}$$

Proof. See Appendix 7.1 □

3.4. Fisher information

In the small deviation regime, we assume that δSNR is a small deviation of the SNR. The new binary hypothesis test is

$$\begin{cases} \mathcal{H}_0 & : \quad \mathbf{y}|\delta\mathrm{SNR} = 0 \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}(0)\right), \\ \mathcal{H}_1 & : \quad \mathbf{y}|\delta\mathrm{SNR} \neq 0 \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{\Sigma}(\delta\mathrm{SNR})\Big) \end{cases}$$

where $\mathbf{\Sigma}(x) = x \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}$. The s-divergence in the small SNR deviation scenario is written as

$$\mu_N(s) = \frac{1-s}{2} \log\det\left[\mathbf{\Sigma}(\delta\mathrm{SNR})\right] - \frac{1}{2} \log\det\left[\mathbf{\Sigma}(\delta\mathrm{SNR} \cdot (1-s))\right]$$

Lemma 7. *The s-divergence in the small deviation regime can be approximated according to*

$$\frac{\mu_N(s)}{N} \overset{\delta\mathrm{SNR} \ll 1}{\approx} (s-1)s \cdot \frac{(\delta\mathrm{SNR})^2}{2} \cdot \frac{\mathrm{J}_F(0)}{N}$$

where the Fisher information [34] is given by

$$\mathrm{J}_F(x) = \frac{1}{2}\mathrm{Tr}((\mathbf{I} + x \cdot \mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{\Phi}^T(\mathbf{I} + x \cdot \mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{\Phi}^T).$$

Proof. See Appendix 7.2 □

According to Lemma 7, the optimal s-value at low SNR is $s^\star \overset{\delta\mathrm{SNR} \ll 1}{=} \frac{1}{2}$. At contrary, the optimal s-value for larger SNR is given by the following lemma.

Lemma 8. *In case of large* SNR, *we have*

$$s^\star \overset{\mathrm{SNR} \gg 1}{\approx} 1 - \frac{1}{\log\mathrm{SNR} + \frac{1}{K}\sum_{n=1}^{K} \log\lambda_n}. \tag{9}$$

where $(\lambda_n)_{n=1,\ldots,N}$ *are the eigenvalues of* $\mathbf{\Phi}\mathbf{\Phi}^T$.

Proof. See Appendix 7.3 □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

4. Computational Information Geometry for classification

4.1. Formulation of the observation vector as a structured linear model

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

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{N}} \tag{10}$$

where $\boldsymbol{\mathcal{N}}$ is the noise tensor where each entry is assumed to be centered *i.i.d.* Gaussian, *i.e.* $[\boldsymbol{\mathcal{N}}]_{n_1,\ldots,n_Q} \sim \mathcal{N}(0,\sigma^2)$ and the noise-free tensor $\boldsymbol{\mathcal{X}}$ follows either CPD or TKD given by definition 2.1.2 and definition 2.1.3, respectively. The vectorization of (10) is given by

$$\mathbf{y}_N = \text{vec}(\mathbf{Y}_{(1)}) = \mathbf{x} + \mathbf{n} \tag{11}$$

where $\mathbf{n} = \text{vec}(\mathbf{N}_{(1)})$ and $\mathbf{x} = \text{vec}(\mathbf{X}_{(1)})$. Note that $\mathbf{Y}_{(1)}$, $\mathbf{N}_{(1)}$ and $\mathbf{X}_{(1)}$ are respectively the first unfolding matrices given by definition 4 of tensors $\boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{N}}$ and $\boldsymbol{\mathcal{X}}$,

1. When tensor $\boldsymbol{\mathcal{X}}$ follows a Q-order CPD with a canonical rank of M, we have

$$\mathbf{x} = \text{vec}\left\{\boldsymbol{\Phi}^{(1)}\mathbf{S}\left(\boldsymbol{\Phi}^{(Q)} \odot \ldots \odot \boldsymbol{\Phi}^{(2)}\right)^T\right\} = \boldsymbol{\Phi}^{\odot}\boldsymbol{s}$$

   where $\boldsymbol{\Phi}^{\odot} = \boldsymbol{\Phi}^{(Q)} \odot \ldots \odot \boldsymbol{\Phi}^{(1)}$ is a $N \times R$ structured matrix and $\boldsymbol{s} = \begin{bmatrix} s_1 & \ldots & s_R \end{bmatrix}^T$ where $s_r \sim \mathcal{N}(0,\sigma_s^2)$, *i.i.d.* and $N = N_1 \cdots N_Q$.
2. When tensor $\boldsymbol{\mathcal{X}}$ follows a Q-order TKD of multilinear rank of $\{M_1,\ldots,M_Q\}$, we have

$$\mathbf{x} = \text{vec}\left\{\boldsymbol{\Phi}^{(1)}\mathbf{S}_{(1)}\left(\boldsymbol{\Phi}^{(Q)} \otimes \ldots \otimes \boldsymbol{\Phi}^{(2)}\right)^T\right\} = \boldsymbol{\Phi}^{\otimes}\text{vec}(\boldsymbol{\mathcal{S}})$$

   where $\boldsymbol{\Phi}^{\otimes} = \boldsymbol{\Phi}^{(Q)} \otimes \ldots \otimes \boldsymbol{\Phi}^{(1)}$ is a $N \times M$ structured matrix with $M = M_1 \cdots M_Q$ and $\text{vec}(\boldsymbol{\mathcal{S}})$ is the vectorization of tensor $\boldsymbol{\mathcal{S}}$ where $s_{m_1,\ldots,m_Q} \sim \mathcal{N}(0,\sigma_s^2)$, *i.i.d.*.

4.2. The CPD case

We recall that in the CPD case, matrix $\boldsymbol{\Phi}^{\odot} = \boldsymbol{\Phi}^{(Q)} \odot \ldots \odot \boldsymbol{\Phi}^{(1)}$ and $(\boldsymbol{\Phi}^{(q)})_{q=1,\ldots,Q}$ are matrices of size $N_q \times R$. In the following, we assume that matrices $\boldsymbol{\Phi}^{(q)}_{q=1,\ldots,Q}$ are random matrices with Gaussian $\mathcal{N}(0,\frac{1}{N_q})$ variate entries. We evaluate the behavior of $\frac{\mu_N(s)}{N}$ when $(N_q)_{q=1,\ldots,Q}$ converge towards $+\infty$ at the same rate and that $\frac{R}{N}$ converges towards a non zero limit.

Result 9. *In the asymptotic regime where $N_1,\ldots,N_Q$ converge towards $+\infty$ at the same rate and where $R \to +\infty$ in such a way that $c_R = \frac{R}{N}$ converges towards a finite constant $c > 0$, it holds that*

$$\frac{\mu_N(s)}{N} \xrightarrow{\text{a.s}} \mu(s) = \frac{1-s}{2}\Psi_c(\text{SNR}) - \frac{1}{2}\Psi_c((1-s)\cdot\text{SNR}) \tag{12}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

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

$$\begin{aligned}\Psi_c(x) = &\log\left(1+\frac{2c}{u(x)+(1-c)}\right)\\ &+c\cdot\log\left(1+\frac{2}{u(x)-(1-c)}\right)\\ &-\frac{4c}{x(u(x)^2-(1-c)^2)}\end{aligned} \tag{13}$$

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 7.4. □

Remark 10. *In [37], the Central Limit Theorem (CLT) for the linear eigenvalue statistics of the tensor version of the sample covariance matrix of type* $\mathbf{\Phi}^{\odot}(\mathbf{\Phi}^{\odot})^T$ *is established, for* $\mathbf{\Phi}^{\odot}=\mathbf{\Phi}^{(2)}\odot\mathbf{\Phi}^{(1)}$, *i.e the tensor order is* $Q=2$.

4.2.1. Small SNR deviation scenario

In this section, we assume that SNR is small. Under this regime, we have the following result:

Result 11. *In the small* SNR *scenario, the Fisher information for CPD is given as*

$$\mu\left(\frac{1}{2}\right)\overset{\text{SNR}\ll 1}{\approx}-\frac{(\text{SNR})^2}{16}\cdot c(1+c).$$

Proof. Using lemma 7, we can notice that

$$\frac{\mathrm{J}_F(0)}{N}=\frac{1}{2}\frac{R}{N}\frac{1}{R}\mathrm{Tr}\left[(\mathbf{\Phi}^{\odot}(\mathbf{\Phi}^{\odot})^T)^2\right]$$

and that

$$\frac{1}{R}\mathrm{Tr}\left[(\mathbf{\Phi}^{\odot}(\mathbf{\Phi}^{\odot})^T)^2\right]$$

converges a. s. towards the second moment of the Marchenko-Pastur distribution which is $1+c$ (see for instance [5]). □

Note that $\mu\left(\frac{1}{2}\right)$ is the error exponent related to the Bhattacharyya divergence.

4.2.2. Large SNR deviation scenario

Result 12. *In case of large* SNR, *the minimizer of Chernoff Information is given by*

$$s^{\star}\overset{\text{SNR}\gg 1}{\approx}1-\frac{1}{\log\text{SNR}-1-\frac{1-c}{c}\log(1-c)}. \tag{14}$$

Proof. It is straightforward to notice that

$$\frac{1}{K}\sum_{n=1}^{K}\log(\lambda_n)\longrightarrow\int_0^{+\infty}\log(\lambda)d\nu_c(\lambda)=-1-\frac{1-c}{c}\log(1-c).$$

The last equality can be obtained as in [50]. Using lemma 8, we get immediately (14). □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

Remark 13. *It is interesting to note that for $c \to 0$ or 1, the optimal s-value follows the same approximated relation given by*

$$s^\star \overset{\text{SNR}\gg 1}{\approx} 1 - \frac{1}{\log \text{SNR}}$$

as long as $\text{SNR} \gg \exp[1]$ *or equivalently a* SNR *in dB much larger than 4 dB.*

Proof. It is straightforward to note that

$$\frac{1-c}{c}\log(1-c) \overset{c\to 1}{\longrightarrow} 0, \text{ and } \frac{1-c}{c}\log(1-c) \overset{c\to 0}{\longrightarrow} -1.$$

Using eq. (14) and condition $\text{SNR} \gg \exp[1]$, the desired result is proved. □

4.2.3. Approximated analytical expressions for $c \ll 1$ and any SNR

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

Result 14. *In this context, the error exponent can be approximated as follows:*

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

Proof. See Appendix 7.5. □

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 $s^\star$ is a global minimizer and is given by zeroing the first-order derivative of the error exponent. This optimal value is given by

$$s^\star \overset{c\ll 1}{\approx} 1 + \frac{1}{\text{SNR}} - \frac{1}{\log(1+\text{SNR})}. \tag{15}$$

We can identify the two following limit scenarios:

- At low SNR, the error exponent associated with the tightest CUB, denoted by $\mu(s^\star)$, coincides with the error exponent associated with the BUB. Indeed, the optimal value in eq. (15) admits a second-order approximation for $c \ll 1$ according to

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

Using Result 9 and the above approximation, the best error exponent at low SNR and for $c \ll 1$ is given by

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

- At contrary for $\text{SNR} \to \infty$, we have $s^\star \to 1$. So, the error exponent associated to BUB cannot be considered as optimal in this regime. Using eq. (15) in Result 14 and assuming that $\frac{\log \text{SNR}}{\text{SNR}} \to 0$, the optimal error exponent for large SNR can be approximated according to

$$\mu\left(s^\star\right) \overset{\text{SNR}\gg 1}{\approx} \frac{c}{2}\left(1 - \log\text{SNR} + \log\log(1+\text{SNR})\right).$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

4.3. The TKD case

In the TKD case, we recall that matrix $\mathbf{\Phi}^{\otimes} = \mathbf{\Phi}^{(Q)} \otimes \ldots \otimes \mathbf{\Phi}^{(1)}$, with $(\boldsymbol{\phi}^{(q)})_{1\leq q\leq Q}$ are $N_q \times M_q$ dimensional matrices. We still assume that matrices $\mathbf{\Phi}^{(q)}_{q=1,\ldots,Q}$ are random matrices with Gaussian $\mathcal{N}(0, \frac{1}{N_q})$ entries.

Result 15. *In the asymptotic regime where $M_q < N_q, 1 \leq q \leq Q$ and M_q, N_q converge towards $+\infty$ at the same rate such that $\frac{M_q}{N_q} \to c_q$, where $0 < c_q < 1$, it holds*

$$\frac{\mu_N(s)}{N} \xrightarrow{\text{a.s}} \mu(s) = c_1 \cdots c_Q \left[\frac{1-s}{2} \int_0^{+\infty} \cdots \int_0^{+\infty} \log(1 + \text{SNR} \cdot \lambda_1 \cdots \lambda_Q) d\nu_{c_1}(\lambda_1) \cdots d\nu_{c_Q}(\lambda_Q) \right.$$
$$\left. - \frac{1}{2} \int_0^{+\infty} \cdots \int_0^{+\infty} \log(1 + (1-s)\text{SNR} \cdot \lambda_1 \cdots \lambda_Q) d\nu_{c_1}(\lambda_1) \cdots d\nu_{c_Q}(\lambda_Q) \right] \quad (16)$$

where ν_{c_q} are Marchenko-Pastur distributions of parameters $(c_q, 1)$ defined as in eq. (1).

Proof. See Appendix 7.6. □

Remark 16. *We can notice that for $Q = 1$, the result 15 is similar to result 9. However, when $Q \geq 2$, the integrals in eq. (16) are not tractable in a closed-form expression. For instance, let $Q = 2$, we consider the integral*

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \log(1 + \text{SNR} \cdot \lambda_1 \lambda_2) \nu_{c_1}(d\lambda_1) \nu_{c_2}(d\lambda_2)$$
$$= \int_{\lambda_{c_1}^-}^{\lambda_{c_1}^+} \int_{\lambda_{c_2}^-}^{\lambda_{c_2}^+} \log(1 + \text{SNR} \cdot \lambda_1 \lambda_2) \frac{\sqrt{(\lambda_1 - \lambda_{c_1}^-)(\lambda_{c_1}^+ - \lambda_1)}}{2\pi c_1 \lambda_1} \frac{\sqrt{(\lambda_2 - \lambda_{c_2}^-)(\lambda_{c_2}^+ - \lambda_2)}}{2\pi c_2 \lambda_2} d\lambda_1 \, d\lambda_2$$

where $\lambda_{c_i}^{\pm} = (1 \pm \sqrt{c_i})^2, i = 1, 2$. We can notice that this integral is characterized by elliptic integral (see e.g [1]). As a consequence, it cannot be expressed in closed-form. However, numerical computations can be exploited to solve efficiently the minimization problem of eq. (7).

4.3.1. Large SNR deviation scenario

Result 17. *In case of large* SNR*, the minimizer of Chernoff Information for TKD is given by*

$$s^\star \overset{\text{SNR}\gg 1}{\approx} 1 - \frac{1}{\log \text{SNR} - Q - \sum_{i=1}^{Q} \frac{1-c_i}{c_i} \log(1-c_i)}. \quad (17)$$

Proof. We have that

$$\frac{1}{M} \sum_{n=1}^{M} \log(\lambda_n) \longrightarrow \sum_{q=1}^{Q} \int_0^{+\infty} \log(\lambda_q) d\nu_{c_q}(\lambda_q)$$
$$= \sum_{q=1}^{Q} \left(-1 - \frac{1-c_q}{c_q} \log(1-c_q) \right)$$
$$= -Q - \sum_{q=1}^{Q} \frac{1-c_q}{c_q} \log(1-c_q).$$

Using lemma 8, we get immediately (17). □

4.3.2. Small SNR deviation scenario

Under this regime, we have the following results

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

Result 18. *For small* SNR *deviation, the Chernoff information for the TKD is given by*

$$\mu\left(\frac{1}{2}\right) \overset{\delta\text{SNR}\ll 1}{\approx} -\frac{(\delta\text{SNR})^2}{16}\prod_{q=1}^{Q} c_q \cdot (1+c_q).$$

Proof. Using lemma 7, we can notice that

$$\frac{\text{J}_F(0)}{N} = \frac{1}{2}\frac{M}{N}\frac{1}{M}\text{Tr}\left[(\mathbf{\Phi}^{\otimes}(\mathbf{\Phi}^{\otimes})^T)^2\right] = \frac{1}{2}\frac{M}{N}\prod_{q=1}^{Q}\frac{\text{Tr}\left[(\mathbf{\Phi}^{(q)}\mathbf{\Phi}^{(q)^T})^2\right]}{M_q}.$$

Each term in the product converges *a.s.* towards the second moment of Marchenko-Pastur distributions ν_{c_q} which are $1+c_q$ and $\frac{M}{N}$ converges to $\prod_{q=1}^{Q} c_q$. This proves the desired result. □

Remark 19. *Contrary to the remark 13, it is interesting to note that for $c_1 = c_2 = ... = c_Q = c$ and $c \to 0$ or* 1*, the optimal s-value follows different approximated relation given by*

$$s^\star \underset{c\to 0}{\overset{\text{SNR}\gg 1}{\approx}} 1 - \frac{1}{\log \text{SNR}}$$

which does not depend on Q, and

$$s^\star \underset{c\to 1}{\overset{\text{SNR}\gg 1}{\approx}} 1 - \frac{1}{\log \text{SNR} - Q}$$

which depends on Q.

In practice, when c is close to 1, we have to carefully check if Q is in the neighbourhood of $\log(\text{SNR})$*. As we can see that, when* $\log \text{SNR} - Q < 0$ *or* $0 < \log \text{SNR} - Q < 1$*, following the above approximation,* $s^\star \notin [0,1]$.

5. Numerical illustrations

In this simulation part, we consider cubic tensors of order $Q = 3$ with $N_1 = 10, N_2 = 20, N_3 = 30, R = 3000$ following a CPD and $M_1 = 100, M_2 = 120, M_3 = 140, N_1 = N_2 = N_3 = 200$ for the TKD, respectively.

Firstly, for the CPD model, in Fig. 4, it is drawn parameter $s^\star$ with respect to the SNR in dB. The parameter $s^\star$ is obtained thanks to three different methods. The first one is based on the brute force/exhaustive computation of the CUB by minimizing the expression in eq. (8) with $\mathbf{\Phi} = \mathbf{\Phi}^{\odot}$. This approach has a very high computational cost especially in our asymptotic context (for a standard computer with Intel Xeon E5-2630 2.3GHz and 32GB RAM, it requires 183 hours to establish 10000 simulations). The second approach is based on the numerical optimization of the closed-form expression of $\mu(s)$ given in Result 14. In this scenario, the drawback in terms of the computational cost is largely mitigated since it consists of a minimization of an univariate regular function. Finally, under the hypothesis that SNR is large, typically > 30 dB, the optimal s-value, $s^\star$, is derived by an analytic expression given by eq. (15). We can check that the the proposed semi-analytic and analytic expressions are in good agreement with the brute-force method for a lowest computational cost. Moreover, we compute the mean square relative error $\frac{1}{L}\sum_{l=1}^{L}(\frac{\widehat{s}_l^\star - s^\star}{s^\star})^2$ where $L = 10000$ the number of samples for Monte-Carlo process and where $\widehat{s}_l^\star = \arg\min_{s\in[0,1]} \mu_{N,l}(s)$ and $s^\star = \arg\min_{s\in[0,1]} \mu(s)$. It turns out that the mean square relative errors are in mean of order -40 dB. We can conclude that the estimator $\widehat{s}^\star$ is a consistent estimator of $s^\star$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

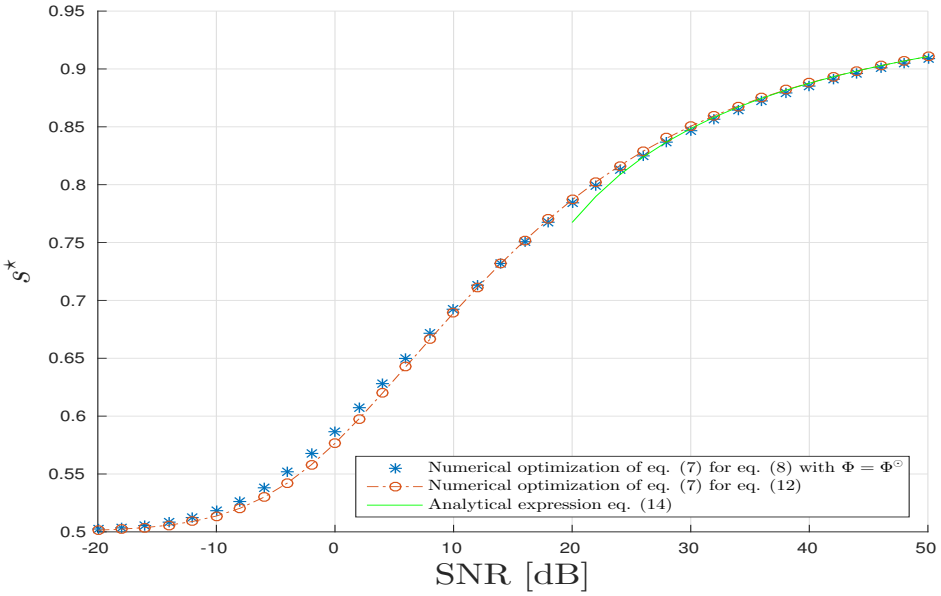

Figure 4. CPD scenario: Optimal s-parameter *vs* SNR in dB

In Fig. 5, we draw various s-divergences: $\mu\left(\frac{1}{2}\right), \mu(s^\star), \frac{1}{N}\mu_N\left(\frac{1}{2}\right), \frac{1}{N}\mu_N(\hat{s})$. We can observe the good agreement with the proposed theoretical results. The s-divergence obtained by fixing $s = \frac{1}{2}$ is accurate only at small SNR but degrades when SNR grows large.

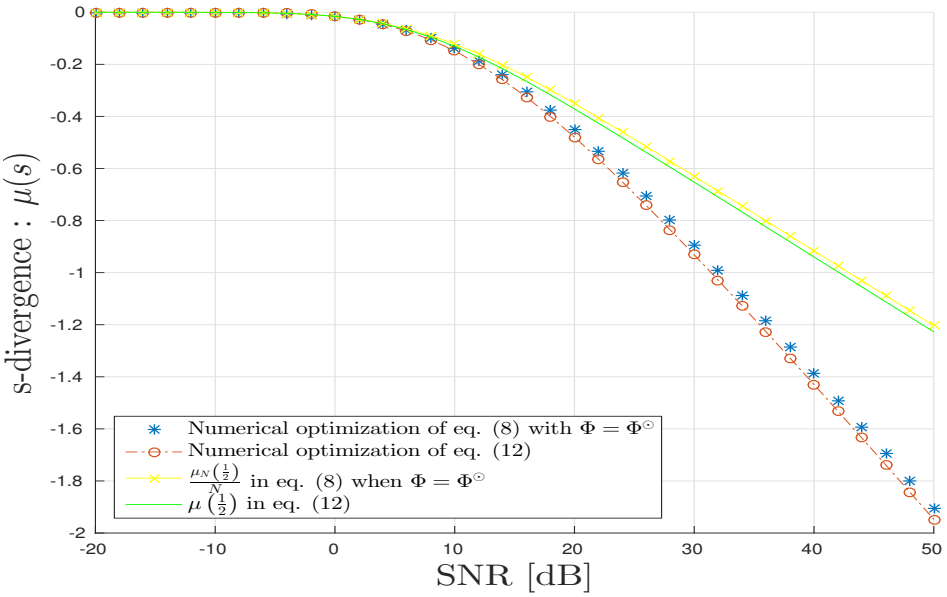

Figure 5. CPD scenario : s-divergence *vs* SNR in dB

In Fig. 6, we fixe SNR $=$ 45 dB and draw $s^\star$ obtained by eq. (14) versus values of $c \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.25, 0.5, 0.75, 0.9, 0.99\}$ and the expression obtained by eq. (15). The two curves approach each other as c goes to zero as predicted by our theoretical analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

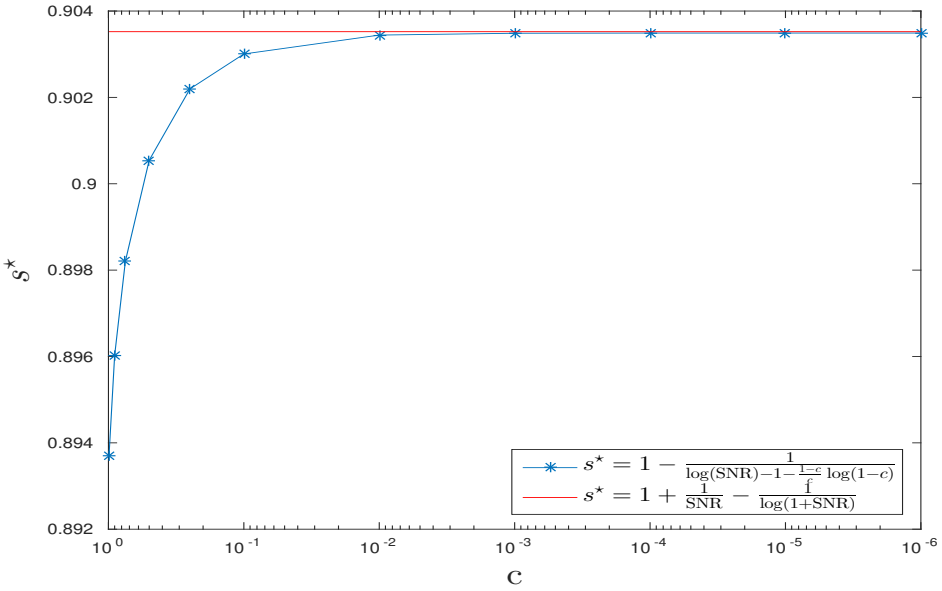

Figure 6. CPD scenario: $s^\star$ *vs* c , SNR = 45 dB

For the TKD scenario, we follow the same methodology as above for CPD, Fig. 7 and Fig. 8 all agree with the analysis provided in section 4.3.

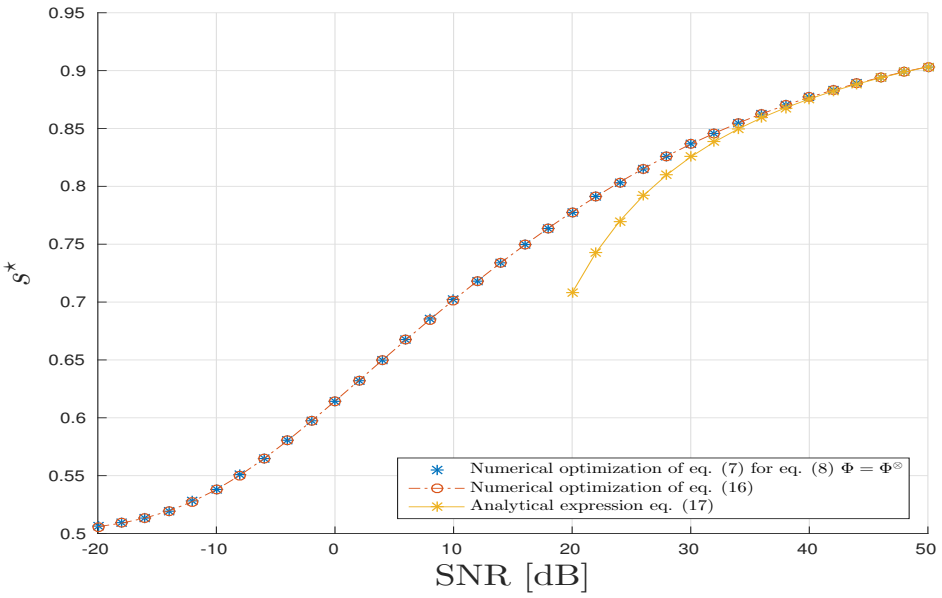

Figure 7. TKD scenario : Optimal *s*-parameter *vs* SNR in dB

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

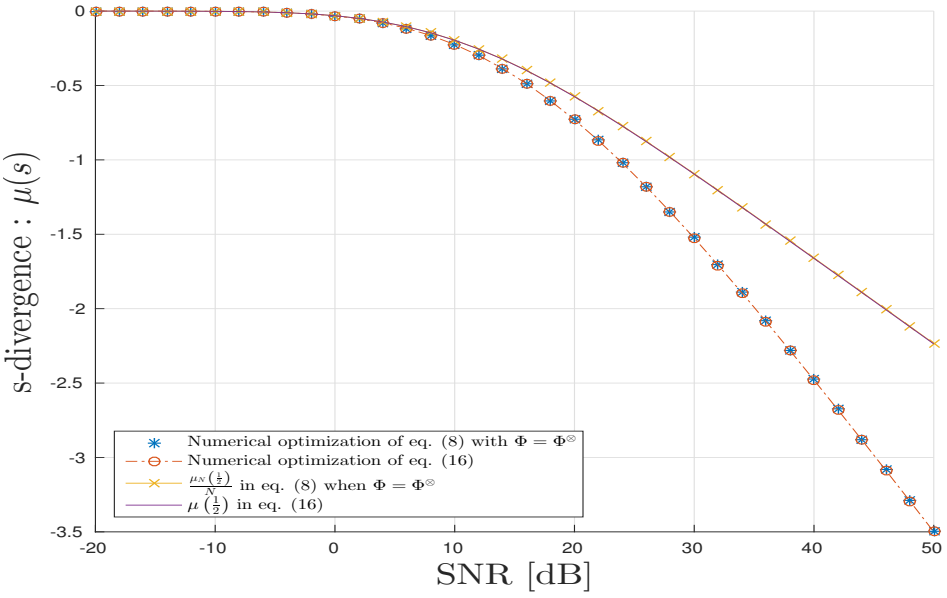

Figure 8. TKD scenario : s-divergence *vs* SNR in dB

For TKD scenario, the mean square relative error is in mean of order -40 dB. So, we check numerically the consistency of the estimator of the optimal s-value.

We can also notice that the convergence of $\frac{\mu_N(s)}{N}$ towards its deterministic equivalent $\mu(s)$ in the case TKD is faster than in the case CPD, since the dimension of matrix $\mathbf{\Phi}^{\otimes}$ is $200.200.200 \times 100.120.140$ ($N = 200^3$) which is much larger than the dimension 6000×3000 of $\mathbf{\Phi}^{\odot}$ ($N = 6000$).

6. Conclusion

In this work, we derived and studied the limit performance in terms of minimal Bayes' error probability for the detection of high-dimensional random tensors using both the tools of Information Geometry (IG) and of Random Matrix Theory (RMT). The main results on Chernoff Bounds and Fisher Information are illustrated by Monte-Carlo simulations that corroborated our theoretical analysis.

For future work, we would like to study the rate of convergence and the fluctuation of the statistics $\frac{\mu_N(s)}{N}$ and $\hat{s}$.

7. Appendix

7.1. Proof of Lemma 6

The s-divergence in eq. (6) for the following binary hypothesis test

$$\begin{cases} \mathcal{H}_0 & : \quad \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0), \\ \mathcal{H}_1 & : \quad \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1) \end{cases}$$

is given by [40]:

$$\mu_N(s) = \frac{1}{2} \log \frac{\det(s\mathbf{\Sigma}_0 + (1-s)\mathbf{\Sigma}_1)}{[\det\mathbf{\Sigma}_0]^s [\det\mathbf{\Sigma}_1]^{1-s}}. \tag{18}$$

Using the expressions of the covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, the numerator in eq. (18) is given by

$$N \log \sigma^2 + \log \det \left(\text{SNR} \cdot (1-s) \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I} \right)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

and the two terms at its numerator are $\log[\det \mathbf{\Sigma}_0]^s = sN \log \sigma^2$ and

$$\log[\det \mathbf{\Sigma}_1]^{1-s} = (1-s)\Big(N \log \sigma^2 + \log\det\Big(\mathrm{SNR} \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\Big)\Big).$$

Using the above expressions, $\mu_N(s)$ is given by eq. (8).

7.2. Proof of Lemma 7

If we note $d\mathbf{\Sigma}(\mathrm{SNR}) = \frac{\partial \mathbf{\Sigma}(x)}{\partial x}\Big|_{x=\mathrm{SNR}}$ then the following expression holds:

$$\mathbf{\Sigma}(\delta\mathrm{SNR}) = \mathbf{\Sigma}(0) + (\delta\mathrm{SNR}) \cdot d\mathbf{\Sigma}(0) = \mathbf{I} + (\delta\mathrm{SNR}) \cdot \mathbf{\Phi}\mathbf{\Phi}^T.$$

Using the above expression, the s-divergence is given by

$$\mu_N(s) = \frac{1-s}{2} \log\det\Big[\mathbf{I} + (\delta\mathrm{SNR}) \cdot \mathbf{\Phi}\mathbf{\Phi}^T\Big] - \frac{1}{2}\log\det\Big[\mathbf{I} + \delta\mathrm{SNR} \cdot (1-s) \cdot \mathbf{\Phi}\mathbf{\Phi}^T\Big]$$

Now, using eq. (8), and the following approximation:

$$\frac{1}{N}\log\det(\mathbf{I} + x\mathbf{A}) = \frac{1}{N}\mathrm{Tr}\log(\mathbf{I} + x\mathbf{A}) \approx x \cdot \frac{1}{N}\mathrm{Tr}\mathbf{A} - \frac{x^2}{2} \cdot \frac{1}{N}\mathrm{Tr}\mathbf{A}^2$$

we obtain

$$\frac{\mu_N(s)}{N} \approx (s-1)s \cdot \frac{(\delta\mathrm{SNR})^2}{2} \cdot \frac{\mathrm{J}_F(0)}{N}$$

where the Fisher information for $\mathbf{y}|\delta\mathrm{SNR} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\delta\mathrm{SNR}))$ is given by [34]:

$$\begin{aligned}
\mathrm{J}_F(\delta\mathrm{SNR}) &= -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{y}|\delta\mathrm{SNR})}{\partial(\delta\mathrm{SNR})^2}\right] \\
&= \frac{1}{2}\mathrm{Tr}\{\mathbf{\Sigma}(\delta\mathrm{SNR})^{-1} d\mathbf{\Sigma}(\delta\mathrm{SNR}) \mathbf{\Sigma}(\delta\mathrm{SNR})^{-1} d\mathbf{\Sigma}(\delta\mathrm{SNR})\} \\
&= \frac{1}{2}\mathrm{Tr}((\mathbf{I} + (\delta\mathrm{SNR}) \cdot \mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{\Phi}^T(\mathbf{I} + \delta\mathrm{SNR}) \cdot \mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{\Phi}^T).
\end{aligned}$$

7.3. Proof of Theorem 8

The first step of the proof is based on the derivation of an alternative expression of $\mu_s(\mathrm{SNR})$ given by eq. (18) involving the inverse of the covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$. Specifically, we have

$$\begin{aligned}
\mu_s(\mathrm{SNR}) &= \frac{1}{2}\log\frac{(\det\mathbf{\Sigma}_0)(\det\mathbf{\Sigma}_1)\det((1-s)\mathbf{\Sigma}_0^{-1} + s\mathbf{\Sigma}_1^{-1})}{[\det\mathbf{\Sigma}_0]^s[\det\mathbf{\Sigma}_1]^{1-s}} \\
&= -\frac{1}{2}\log\frac{\det\Big([(1-s)\mathbf{\Sigma}_0^{-1} + s\mathbf{\Sigma}_1^{-1}]^{-1}\Big)}{[\det\mathbf{\Sigma}_0]^{1-s}[\det\mathbf{\Sigma}_1]^s}.
\end{aligned} \tag{19}$$

The second step is to derive a closed-form expression in the high SNR regime using the following the approximation (see [7] for instance): $\Big(x \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\Big)^{-1} \overset{x \gg 1}{\approx} \mathbf{\Pi}_{\mathbf{\Phi}}^{\perp} = \mathbf{I}_N - \mathbf{\Phi}\mathbf{\Phi}^{\dagger}$ where $\mathbf{\Pi}_{\mathbf{\Phi}}^{\perp}$ is an orthogonal projector such as $\mathbf{\Pi}_{\mathbf{\Phi}}^{\perp}\mathbf{\Phi} = \mathbf{0}$ and $\mathbf{\Phi}^{\dagger} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T$. The numerator in eq. (19) is given by

$$\begin{aligned}
\Big[(1-s)\mathbf{\Sigma}_0^{-1} + s\mathbf{\Sigma}_1^{-1}\Big]^{-1} &\overset{\mathrm{SNR} \gg 1}{\approx} \sigma^2\Big(\mathbf{I}_N - s\mathbf{I}_N + s\mathbf{\Pi}_{\mathbf{\Phi}}^{\perp}\Big)^{-1} \\
&= \sigma^2\Big(\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^{\dagger}\Big)^{-1}.
\end{aligned}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

As $s\mathbf{\Phi}\mathbf{\Phi}^\dagger$ is a rank-K projector matrix scaled by factor $s > 0$, its eigen-spectrum is given by $\{\underbrace{s,\ldots,s}_{K},\underbrace{0,\ldots,0}_{N-K}\}$. In addition, as the rank-N identity matrix and the scaled projector $s\mathbf{\Phi}\mathbf{\Phi}^\dagger$ can be diagonalized in the same orthonormal basis matrix, the n-th eigenvalue of the inverse of matrix $\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^\dagger$ is given by

$$\begin{aligned}\lambda_n\left\{\left(\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^\dagger\right)^{-1}\right\} &= \frac{1}{\lambda_n\left\{\mathbf{I}_N\right\} - s\lambda_n\left\{\mathbf{\Phi}\mathbf{\Phi}^\dagger\right\}} \\ &= \begin{cases} \frac{1}{1-s}, & 1 \le n \le K, \\ 1, & K+1 \le n \le N \end{cases}\end{aligned}$$

with $s \in (0,1)$. Using the above property, we obtain

$$\begin{aligned}\log\det\left([\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^\dagger]^{-1}\right) &= \log\prod_{n=1}^{N}\lambda_n\left\{\left(\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^\dagger\right)^{-1}\right\} \\ &= -K\log(1-s).\end{aligned}$$

In addition, we have

$$\log\det\left(\mathrm{SNR}\cdot\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \overset{\mathrm{SNR}\gg 1}{\approx} \mathrm{Tr}\log\left(\mathrm{SNR}\cdot\mathbf{\Phi}^T\mathbf{\Phi}\right) = K\cdot\log\mathrm{SNR} + \sum_{n=1}^{K}\log\lambda_n$$

Finally, thanks to eq. (19), we have

$$\frac{\mu_s(\mathrm{SNR})}{N} \overset{\mathrm{SNR}\gg 1}{\approx} \frac{1}{2}\frac{K}{N}\left(\log(1-s) + s\cdot\log\mathrm{SNR} + \frac{s}{K}\sum_{n=1}^{K}\log\lambda_n\right)$$

Finally, to obtain $s^\star$ in eq. (9), we solve $\frac{\partial\mu_s(\mathrm{SNR})}{\partial s} = 0$.

7.4. Proof of Result 9

Large random matrix theory allows to evaluate the asymptotic behavior of $\frac{\mu_N(s)}{N}$ when $N_q \to +\infty$ for each $q = 1,\ldots,Q$, $R \to +\infty$ in such a way that $\frac{R^{1/q}}{N_q}$ converge towards a non zero constant for each $q = 1,\ldots,Q$. In other words, $N_1,\ldots,N_Q$ 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 $\mathbf{\Phi}^{\odot}(\mathbf{\Phi}^{\odot})^T$ converges towards a relevant Marcenko-Pastur distribution. More precisely, we define the Marcenko-Pastur distribution $\nu_c(d\lambda)$ as the probability distribution given by

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

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

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

$$t_c(-\rho) = \frac{2}{\rho - (1-c) + \sqrt{(\rho + \lambda_c^-)(\rho + \lambda_c^+)}} \tag{20}$$

It was established for the first time in [38] that if $\mathbf{X}$ represents a $K \times P$ random matrix with zero mean and $\frac{1}{K}$ variance i.i.d. entries, and if $(\lambda_k)_{k=1,\ldots,K}$ 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 $\frac{1}{K}\sum_{k=1}^{K}\delta(\lambda - \lambda_k)$ converges weakly almost surely towards ν_c in the asymptotic regime where $K \to +\infty$, $P \to +\infty$, $\frac{P}{K} \to c$. In particular, for each continuous function $f(\lambda)$, it holds that

$$\frac{1}{K}\sum_{k=1}^{K} f(\lambda_k) \xrightarrow{\text{a.s}} \int_{\mathbb{R}^+} f(\lambda)\, \nu_c(d\lambda). \tag{21}$$

In practice, this result means that if K and P 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 ν_c.
The columns $(\boldsymbol{\phi}_r)_{r=1,\ldots,R}$ of $\boldsymbol{\Phi}^{\odot}$ are vectors $(\boldsymbol{\phi}_r^{(Q)} \otimes \ldots \otimes \boldsymbol{\phi}_r^{(1)})_{r=1,\ldots,R}$. These vectors are mutually independent, identically distributed, and satisfy $\mathbb{E}(\boldsymbol{\phi}_r \boldsymbol{\phi}_r^T) = \frac{I_N}{N}$. However, the elements of $\boldsymbol{\Phi}^{\odot}$ are not mutually independent because the components of each column $\boldsymbol{\phi}_r$ are not independent. In the asymptotic regime considered in this paper, the results of [44] (see also [3]) allow to establish that the empirical eigenvalue distribution of $\boldsymbol{\Phi}^{\odot}(\boldsymbol{\Phi}^{\odot})^T$ still converges almost surely towards ν_c, where we recall that $\frac{R}{N} \to c$. Using (21) for $f(\lambda) = \log(1 + \lambda/\rho)$ as well as a well-known formula that allows to express $\int_{\mathbb{R}^+} \log(1 + \lambda/\rho)\, \nu_c(d\lambda)$ in terms of $t_c(-\rho)$ given by (20) (see e.g. [50]), we obtain the result.

7.5. Proof of Result 14

We have $u(x) \overset{c \ll 1}{\approx} \frac{1}{x} + \sqrt{(\frac{1}{x}+1)^2} = \frac{2}{x} + 1$ and $u(x) + (1-c) \overset{c \ll 1}{\approx} 2\left(\frac{1}{x}+1\right)$, $u(x) - (1-c) \overset{c \ll 1}{\approx} \frac{2}{x}$, $u(x)^2 - (1-c)^2 \overset{c \ll 1}{\approx} \frac{4}{x}\left(\frac{1}{x}+1\right)$. Using the above first-order approximations, eq. (13) is

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

Using the above approximation and eq. (12), we obtain Result 14.

7.6. Proof of Result 15

We first denote $\lambda_1^{(q)} \geq \lambda_2^{(q)} \geq \ldots \geq \lambda_{n_q}^{(q)} \geq \ldots \geq \lambda_{N_q}^{(q)}$ the eigenvalues of $\boldsymbol{\Phi}^{(q)}(\boldsymbol{\Phi}^{(q)})^T$, $1 \leq n_q \leq N_q$, for $1 \leq q \leq Q$. We can notice that the eigenvalues of $\boldsymbol{\Phi}^{\otimes}(\boldsymbol{\Phi}^{\otimes})^T$ are $\lambda_{n_1}^{(1)} \cdots \lambda_{n_Q}^{(Q)}$. Moreover, in the asymptotic regime, where $M_q \to +\infty$, $N_q \to +\infty$ such that $\frac{M_q}{N_q} \to c_q$, $0 < c_q < 1$, for all $1 \leq q \leq Q$, we have that $\lambda_{n_q}^{(q)} = 0$ if $M_q + 1 \leq n_q \leq N_q$ and the empirical distribution of the eigenvalues $(\lambda_{n_q}^{(q)})_{1 \leq n_q \leq M_q}$ behaves as Marchenko-Pastur distributions ν_{c_q} of parameters $(c_q, 1)$. Recalling that $M = M_1 \ldots M_Q$, $N = N_1 \ldots N_Q$, we obtain immediately that

$$\begin{aligned}\frac{1}{N}\log\det\left(\mathrm{SNR}\cdot\boldsymbol{\Phi}^{\otimes}(\boldsymbol{\Phi}^{\otimes})^T + \mathbf{I}\right) &= \frac{1}{N}\sum_{n_1=1}^{N_1}\ldots\sum_{n_q=1}^{N_Q}\log\left(\mathrm{SNR}\cdot\lambda_{n_1}^{(1)}\cdots\lambda_{n_Q}^{(Q)}+1\right)\\ &= \frac{M}{N}\frac{1}{M}\sum_{n_1=1}^{M_1}\ldots\sum_{n_q=1}^{M_Q}\log\left(\mathrm{SNR}\cdot\lambda_{n_1}^{(1)}\cdots\lambda_{n_Q}^{(Q)}+1\right)\end{aligned}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0008.v1

and that

$$\frac{1}{M}\sum_{n_1=1}^{M_1}...\sum_{n_q=1}^{M_Q}\log\left(\text{SNR}\cdot\lambda_{n_1}^{(1)}\cdots\lambda_{n_Q}^{(Q)}+1\right)\xrightarrow{a.s}\int_0^{+\infty}...\int_0^{+\infty}\log(1+\text{SNR}\cdot\lambda_1...\lambda_Q)d\nu_{c_1}(\lambda_1)...d\nu_{c_Q}(\lambda_Q)$$

Similarly, we have that

$$\frac{1}{M}\log\det\left(\text{SNR}\cdot(1-s)\mathbf{\Phi}^{\otimes}(\mathbf{\Phi}^{\otimes})^T+\mathbf{I}\right)\xrightarrow{a.s}\int_0^{+\infty}...\int_0^{+\infty}\log(1+\text{SNR}\cdot(1-s)\lambda_1...\lambda_Q)d\nu_{c_1}(\lambda_1)...d\nu_{c_Q}(\lambda_Q)$$

We obtain easily result 15.

Acknowledgments: The authors would like to thank Professor Philippe Loubaton (UPEM, France) for the fruitful discussions.

References

1. M. Abramowitz, I. A. Stegun (Eds.). "Elliptic Integrals." Ch. 17 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing*, New York Dover Publications, pp. 587-607, 1972.
2. S. M. Ali, S. D. Silvey, *A General Class of Coefficients of Divergence of One Distribution from Another*, Journal of the Royal Statistical Society. Series B (Methodological) Vol. 28, No. 1 (1966), pp. 131-142
3. 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).
4. R. Badeau, G. Richard and B. David, *Fast and stable YAST algorithm for principal and minor subspace tracking*. IEEE J SP, IEEE, 2008, 56 (8), pp.3437-3446.
5. Z. D. Bai, J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics, 2nd edition, 2010.
6. J. Baik, J. Silverstein, *Eigenvalues of large sample covariance matrices of spiked population models*, Journal of Multivariate Analysis, Volume 97, Issue 6, July 2006, Pages 1382-1408
7. R. T. Behrens, L. L. Scharf, *Signal processing applications of oblique projection operators*. *IEEE Transactions on Signal Processing* , Vol 42, pp 1413-1424, 1994.
8. O. Besson , L.L. Scharf, "CFAR matched direction detector", *IEEE Transactions on Signal Processing* , Volume: 54, Issue: 7, July 2006.
9. P. Bianchi, M. Debbah, M. Maida, and J. Najim, *Performance of Statistical Tests for Source Detection using Random Matrix Theory*, IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, April 2011.
10. M. Boizard, G. Ginolhac, F. Pascal, and P. Forster, *Low-rank filter and detector for multidimensional data based on an alternative unfolding HOSVD: application to polarimetric STAP* . EURASIP Journal on Advances in Signal Processing, SpringerOpen, 2014, 2014, pp.119.
11. G. Bouleux and R .Boyer, *Sparse-Based Estimation Performance for Partially Known Overcomplete Large-Systems*, Signal Processing, Volume 139, October 2017, pp. 70-74.
12. R. Boyer and R. Badeau. *Adaptive multilinear SVD for structured tensors* . IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'06), 2006, Toulouse, France. 2006.
13. R. Boyer and C. Delpha *Relative-entropy based beamforming for secret key transmission*, IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012.
14. R. Boyer, R. Couillet, B-H. Fleury, and P.Larzabal, *Large-System Estimation Performance in Noisy Compressed Sensing with Random Support - a Bayesian Analysis*, IEEE Transactions on Signal Processing, Volume 64, No. 21, 2016, pp. 5525-5535.
15. R. Boyer and P. Loubaton, *Large deviation analysis of the cpd detection problem based on random tensor theory*, European Association for Signal Processing (EUSIPCO), 2017.
16. 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.
17. 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.

18. S. P. Chepuri and G. Leus, *Sparse sensing for distributed Gaussian detection*,IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.
19. H. Chernoff, *A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations*, Ann. Math. Statist. Volume 23, Number 4 (1952), 493-507.
20. 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.
21. P. Comon, J. T. Berge, L. De Lathauwer and J. Castaing, *Generic and Typical Ranks of Multi-Way Arrays*, Linear Algebra and its Applications, Elsevier, 2009, 430 (11), pp.2997-3007.
22. P. Comon, X. Luciani and A. L. F. de Almeida, *Tensor decompositions, alternating least squares and other tales*, Journal of Chemometrics, 2009.
23. P. Comon , *Tensors : A brief introduction*, IEEE Signal Processing Magazine , Volume: 31, Issue: 3, May 2014.
24. T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012
25. L. De Lathauwer, B. D. Moor, and J. Vandewalle, *A Multilinear Singular Value Decomposition*, SIAM Journal on Matrix Analysis and Applications, 2000, Vol. 21, No. 4 : pp. 1253-1278
26. L. De Lathauwer, *A survey of tensor methods*, IEEE International Symposium on Circuits and Systems, 2009. ISCAS 2009.
27. N. Duy Tran, R. Boyer, S. Marcos, and P. Larzabal, *Angular resolution limit for array processing: Estimation and information theory approaches*, Proceedings of the 20th European Signal Processing Conference (EUSIPCO), 2012.
28. C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, September 1936, Volume 1, Issue 3, pp 211–218.
29. V. L. Girko, *Theory of random determinants*, Kluwer Academic Publishers,1990.
30. J. H. D. M. Goulart, M. Boizard, R. Boyer, G. Favier and P. Comon, *Tensor CP Decomposition with Structured Factor Matrices: Algorithms and Performance*, IEEE Journal of Selected Topics in Signal Processing, IEEE, 2016, 10 (4), pp.757-769.
31. E. Grossi and M. Lops, *Space-time code design for MIMO detection based on Kullback-Leibler divergence*, IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3989-4004, Jun. 2012.
32. T. Kailath, *The Divergence and Bhattacharyya Distance Measures in Signal Selection*. IEEE Transactions on Communication Technology, 15, 52-60, 1967.
33. G. Katz, P. Piantanida, R. Couillet and M. Debbah, *Joint estimation and detection against independence*, Annual Conference on Communication Control and Computing (Allerton), pp. 1220-1227 Sept 2014.
34. S. M. Kay, *Fundamentals of statistical signal processing, Volume II: Detection theory*, PTR Prentice-Hall,Englewood Cliffs, NJ, 1993.
35. Y. Lee, Y. Sung, *Generalized Chernoff Information for Mismatched Bayesian Detection and Its Application to Energy Detection*, IEEE Signal Processing Letters, Volume: 19, Issue: 11, Nov. 2012.
36. P. Loubaton and P. Vallet, *Almost Sure Localization of the Eigenvalues in a Gaussian Information Plus Noise Model. Application to the Spiked Models*. Electron. J. Probab. Volume 16 (2011), paper no. 70, 1934-1959.
37. A. Lytova, *Central Limit Theorem for Linear Eigenvalue Statistics for a Tensor Product Version of Sample Covariance Matrices*, Journal of Theoretical Probability, 2017.
38. 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.
39. X.Mestre, *Improved Estimation of Eigenvalues and Eigenvectors of Covariance Matrices Using Their Sample Estimates*, IEEE Transactions on Information Theory, Volume: 54, Issue: 11, Nov. 2008.
40. F. Nielsen, *Chernoff information of exponential families*. CoRR 2011, abs/1102.2684.
41. F. Nielsen, *An information-geometric characterization of Chernoff information*, IEEE Signal Processing Letters, vol. 20, no. 3, pp.269-272, 2013.
42. F. Nielsen, *Hypothesis testing, information divergence and computational geometry*, Geometric Science of Information, Springer, 2013, pp. 241-248
43. V. Ollier , R. Boyer , M. N. El Korso and P. Larzabal, *Bayesian Lower Bounds for Dense or Sparse (Outlier) Noise in the RMT Framework*, IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 16), 2016. in special session.

44. 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), pp: 11-29.
45. S. Sen, A. Nehorai, *Sparsity-Based Multi-Target Tracking Using OFDM Radar*, IEEE Transactions on Signal Processing, Volume: 59, Issue: 4, April 2011 .
46. J. W. Silverstein, P. L. Combettes, *Signal detection via spectral theory of large dimensional random matrices*. IEEE Transactions on Signal Processing , Volume. 40. No. 8. August 1992.
47. S. Sinanovic, D. H. Johnson, *Toward a theory of information processing*, Signal Processing, vol. 87, no. 6, pp. 1326-1344, 2007.
48. G. Tang, A. Nehorai, *Performance Analysis for Sparse Support Recovery*, IEEE Transactions on Information Theory, Volume: 56, Issue: 3, March 2010.
49. L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika (1966) 31: 279. https://doi.org/10.1007/BF02289464.
50. A. M. Tulino, S. Verdu, *Random Matrix Theory and Wireless Communications*, Hanover, MA, USA: Now Publishers Inc., Jun. 2004, vol. 1, no. 1.
51. D. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math. 104 (1991) 201.
52. E. P. Wigner, *On the statistical distribution of the widths and spacings of nuclear resonance levels*, Proc. Cam. Phil. Soc. 47 (1951) 790.
53. E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions* Ann. Math. 62 (1955) 548.
54. J. Wishart, *Generalized product moment distribution in samples*, Biometrika 20A (1928) 32.