Computational Information Geometry for Binary Classification of High-Dimensional Random Tensors [†]

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

[1] Laboratory of Signals and Systems (L2S), Department of Signals and Statistics, University of Paris-Sud, 91400 Orsay, France; giathuy.pham@l2s.centralesupelec.fr (G.-T.P.); remy.boyer@l2s.centralesupelec.fr (R.B.)
[2] Computer Science Department LIX, École Polytechnique, 91120 Palaiseau, France
[3] Sony Computer Science Laboratories, Tokyo 141-0022, Japan
[*] Correspondence: Frank.Nielsen@acm.org; Tel.: +81-3-5448-4380
[†] The results presented in this work have been partially published in the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017 and the 2017 25th European Association for Signal Processing (EUSIPCO), Kos, Greece, 28 August–2 September 2017.

Abstract: Evaluating the performance of Bayesian classification in a high-dimensional random tensor is a fundamental problem, usually difficult and under-studied. In this work, we consider two Signal to Noise Ratio (SNR)-based binary classification 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 classification of the random entries (coefficients) of the core tensor in the CPD/TKD is hard to study since the exact derivation of the minimal Bayes' error probability is mathematically intractable. To circumvent this 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 $s^\star = 1/2$) 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 tools, a simple analytical expression of $s^\star$ is provided with respect to the Signal to Noise Ratio (SNR) in the two schemes considered. This work shows that the BUB is the tightest bound at low SNRs. However, for higher SNRs, the latest property is no longer true.

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" binary classification problem is a challenging research topic, see for instance [1–7]. 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 [3] since an exact calculation of the minimal Bayes' error probability $P_e^{(N)}$ is often intractable [3,8]. To circumvent this problem, it is standard to exploit well-known bounds on $P_e^{(N)}$ based on information theory [9–13]. In particular, the Chernoff information [14,15] is asymptotically (in N) relied on the exponential rate of $P_e^{(N)}$. It turns out that the Chernoff information is very useful in many practically important problems as for instance, distributed sparse detection [16], sparse support recovery [17], energy detection [18], multi-input and multi-output (MIMO) radar processing [19,20], network secrecy [21], angular resolution limit in array processing [22], detection performance for informed communication systems [23], 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 solving numerically an optimization problem [24] 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 [13] 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 [24].

The tensor decomposition theory is a timely and prominent research topic [25,26]. Confronting the problem of extracting useful information from a massive and multidimensional volume of measurements, it is shown that tensors are extremely relevant. In the standard literature, two main families of tensor decomposition are prominent, namely the Canonical Polyadic Decomposition (CPD) [26] and the Tucker decomposition (TKD)/HOSVD (High-Order SVD) [27,28]. 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 [26] and involves only a reduced number of free parameters due to the constraint of minimality on R. Unfortunately, unlike the matrix case, the set of tensors with fixed (tensorial) rank is not close [29,30]. 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 [31]. Note that this problem can sometimes be avoided by exploiting some natural hidden structures in the physical model [32]. The TKD [28] and the HOSVD [27] are two popular decompositions being an alternative to the CPD. Under this circumstance, alternative definition of rank is required, since the tensorial rank based on CPD scenario is no longer appropriate. In particular, stardard definition of 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. Following the Eckart-Young theorem at each mode level [33], this construction is non-iterative, optimal and practical. In real-time computation [34] or adaptively computation [35], it is shown that this approach is suitable. However, in general, the low (multilinear) rank tensor based on this procedure is suboptimal [27]. More precisely, for tensors of order strictly greater than two, a generalization of the Eckart-Young theorem does not exist.

The classification performance of a multilinear tensor following the CPD and TKD can be derived and studied. It is interesting to note that the classification theory for tensors is very under studied. Based on our knowledge on the topic, only the publication [36] 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 precisely, 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. A standard approach for zero-mean independent Gaussian core and noise tensors, is to define the Signal to Noise Ratio by 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 classification can be described in the following way:

Under the null hypothesis $\mathcal{H}_0$, SNR $= 0$, meaning that the observed tensor contains only noise. Conversely, the alternative hypothesis $\mathcal{H}_1$ is based on SNR $\neq 0$, meaning that there exists a multilinear signal of interest. First note that there exists a lack of contribution dealing with classification performance for tensors. Since the exact derivation of the error probability is intractable, the performance of the classification of the core tensor random entries is hard to evaluate. To circumvent this audible 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 error exponent can be minimized at $s^\star$, which corresponds to the reachable tightest CUB. 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, with respect to different Signal to Noise Ratios (SNR), we provide simple analytical expressions of $s^\star$, thanks to the so-called Random Matrix Theory (RMT). 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 [37–39] 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 attracted both mathematicians and physicists since they were first introduced in mathematical statistics by Wishart in 1928 [40]. When Wigner [41] introduced the concept of statistical distribution of nuclear energy levels, the subject has started to earn prominence. However, it took until 1955 before Wigner [42] introduced ensembles of random matrices. Since then, many important results in RMT were developed and analyzed, see for instance [43–46] and the references therein. In the last two decades, research on RMT has been constantly published.

Finally, let us underline that many arguments of this paper differ from the works presented in [47,48]. In [47], we tackled the problem of detection using Chernoff Upper Bound in data of type matrix in the double asymptotic regime. In [48], we established the detection problem in tensor data by analyzing the Chernoff Upper Bound. In [48], we assumed that the tensor follows the Canonical Polyadic Decomposition (CPD), we gave some analysis of Chernoff Upper Bound when the rank of the tensor is much smaller than the dimensions of the tensor. Since [47,48] are conference papers, some proofs have been omitted due to limited space. Therefore, this full paper may share the ideas in [47,48] on Information Geometry (s-divergence, Chernoff Upper Bound, Fisher Information, etc.), but completes [48] in a more general asymptotic regime. Moreover, in this work, we give new analysis in both scenarios (SNR small and large) whereas [48] did not, and the important and difficult new tensor scenario of the Tucker decomposition is considered. This is in our view the main difference because the CPD is a particular case of the more general decomposition of TucKer. Indeed, in the CPD, the core tensor is assumed to be diagonal.

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.

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 $\text{rank}\{\mathbf{X} \otimes \mathbf{Y}\} = \text{rank}\{\mathbf{X}\} \times \text{rank}\{\mathbf{Y}\}$.

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 $1 \leq k \leq K$.

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

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

where $\circ$ is the outer product [25], $\boldsymbol{\phi}_r^{(q)} \in \mathbb{R}^{N_q \times 1}$ and s_r is a real scalar.

An equivalent formulation using the q-mode product defined in Definition 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 $\mathcal{X}$ is given by

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

where $S = \text{diag}(s)$ with $s = [s_1, ..., s_R]^T$ and $\odot$ stands for the Khatri-Rao product [25].

2.1.3. Tucker Decomposition (TKD)

The Tucker tensor model of order Q is defined according to

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

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 $\mathcal{X}$ is similar to CPD case, however the *q*-mode unfolding matrix for tensor $\mathcal{X}$ is slightly different

$$X_{(q)} = \mathbf{\Phi}^{(q)} S_{(q)} \left(\mathbf{\Phi}^{(Q)} \otimes ... \otimes \mathbf{\Phi}^{(q+1)} \otimes \mathbf{\Phi}^{(q-1)} ... \otimes \mathbf{\Phi}^{(1)} \right)^T$$

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

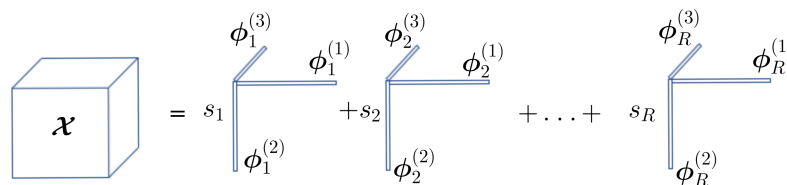**Figure 1.** Canonical Polyadic Decomposition (CPD).

Following the definitions, we note that the CPD and TKD scenarios imply that vector **x** in Equation (11) is related either to the structured linear system $\mathbf{\Phi}^{\odot} = \mathbf{\Phi}^{(Q)} \odot ... \odot \mathbf{\Phi}^{(q+1)} \odot \mathbf{\Phi}^{(q-1)} \odot ... \odot \mathbf{\Phi}^{(1)}$ or $\mathbf{\Phi}^{\otimes} = \mathbf{\Phi}^{(Q)} \otimes ... \otimes \mathbf{\Phi}^{(q+1)} \otimes \mathbf{\Phi}^{(q-1)} ... \otimes \mathbf{\Phi}^{(1)}$.

2.2. The Marchenko-Pastur Distribution

The Marchenko-Pastur distribution was introduced half a century ago [45] 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 $\mathbb{E}(\boldsymbol{v}_n \boldsymbol{v}_n^T) = \sigma^2 I_M$. 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

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

where matrix W_N is defined by $W_N = \frac{1}{\sqrt{N}} [\boldsymbol{v}_1, ..., \boldsymbol{v}_N]$. W_N is thus a Gaussian matrix with independent identically distributed $\mathcal{N}(0, \frac{\sigma^2}{N})$ entries. When $N \rightarrow +\infty$ while M remains fixed, matrix $W_N W_N^T$

converges towards $\sigma^2 I_M$ 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 $\left\| W_N W_N^T - \sigma^2 I_M \right\|$ 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 $W_N W_N^T$ 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
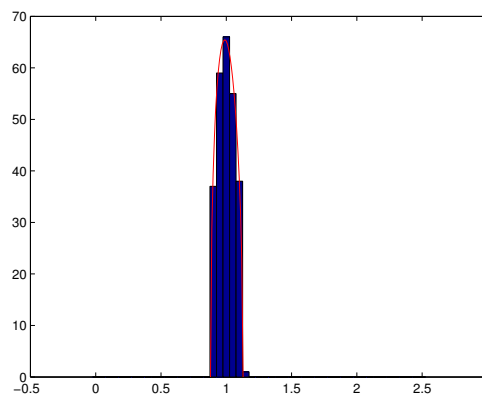
$$\nu_{c,\sigma^2}(d\lambda) = \delta_0[1 - \frac{1}{c}]_+ + \frac{\sqrt{(\lambda - \lambda^-)(\lambda^+ - \lambda)}}{2\sigma^2 c \pi \lambda} \mathbb{1}_{[\lambda^-,\lambda^+]}(\lambda) d\lambda \tag{1}$$

with $\lambda^- = \sigma^2(1 - \sqrt{c})^2$ and $\lambda^+ = \sigma^2(1 + \sqrt{c})^2$. Then, the following result holds.
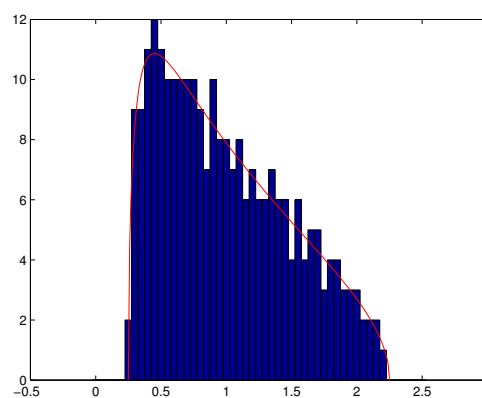
Theorem 1 ([45]). *The empirical eigenvalue value distribution $\hat{\nu}_N$ converges weakly almost surely towards ν_{c,σ^2} when both M and N converge towards $+\infty$ in such a way that $c_N = \frac{M}{N}$ converges towards $c > 0$. Moreover, it holds that*

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

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

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

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

We also observe that Theorem 1 remains valid if W_N is not necessarily a Gaussian matrix whose i.i.d. elements have a finite fourth order moment (see e.g., [43]). Theorem 1 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 $[\lambda^-, \lambda^+]$. See Figures 2 and 3.

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

3.1. Formulation Based on a SNR-Type Criterion

We denote by $\text{SNR} = \sigma_{\mathbf{s}}^2/\sigma^2$ and $p_i(\cdot) = p(\cdot|\mathcal{H}_i)$ with $i \in \{0,1\}$. The binary classification of the random signal based on the equi-probable binary hypothesis test, $\mathbf{s}$, is

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

where $\boldsymbol{\Sigma}_0 = \sigma^2 \mathbf{I}_N$ and $\boldsymbol{\Sigma}_1 = \sigma^2 \left(\text{SNR} \times \boldsymbol{\Phi}\boldsymbol{\Phi}^T + \mathbf{I}_N\right)$. The null hypothesis data-space ($\mathcal{H}_0$) is defined as $\mathcal{X}_0 = \mathcal{X} \setminus \mathcal{X}_1$ 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 alternative hypothesis ($\mathcal{H}_1$) data-space. Following the above expression, the log-likelihood ratio test $\Lambda(\mathbf{y}_N)$ and the binary classification threshold τ' are given by

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

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

where $\det(\cdot)$ and $\log(\cdot)$ are respectively the determinant and the natural logarithm.

3.2. The Expected Log-likelihood Ratio in Geometry Perspective

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

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

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-Leibler Divergence (KLD) [10]. The expected log-likelihood ratio test admits to a simple geometric characterization based on the difference of two KLDs [8]. However, it is often difficult to evaluate the performance of the test via the minimal Bayes' error probability $P_e^{(N)}$, since its expression cannot be determined analytically in closed-form [3,8].

The minimal Bayes' error probability conditionally to vector $\mathbf{y}_N$ is defined as

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

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

3.3. CUB

According to [24], the relation between the Chernoff Upper Bound and the (average) minimal Bayes' error probability $P_e^{(N)} = \mathbb{E}\Pr(\text{Error}|\mathbf{y}_N)$ is given by

$$P_e^{(N)} \leq \frac{1}{2} \times \exp[-\tilde{\mu}_N(s)] \tag{5}$$

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

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

in which $M_X(t) = \mathbb{E}\exp[t \times X]$ is the moment generating function (mgf) of variable X. The error exponent, denoted by $\tilde{\mu}(s)$, 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 [13]

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

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

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

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

$$P_e^{(N)} \leq \frac{1}{2} \times \exp[-\tilde{\mu}_N(s^\star)] \leq \frac{1}{2} \times \exp[-\tilde{\mu}_N(1/2)].$$

Lemma 1. *The log-moment generating function given by Equation (6) for test of Equation (4) is given by*

$$\tilde{\mu}_N(s) = -\frac{1-s}{2}\log\det\left(\text{SNR} \times \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \tag{8}$$
$$+ \frac{1}{2}\log\det\left(\text{SNR} \times (1-s)\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right).$$

Proof. See Appendix A. □

From now on, to simplify the presentation and the numerical results later on, we denote by

$$\mu_N(s) = -\tilde{\mu}_N(s)$$
$$\mu(s) = -\tilde{\mu}(s)$$

for all $s \in [0,1]$, the opposites of the log-moment generating function and its limit.

Remark 1. *The functions $\mu_N(s), \mu(s)$ are negative, since the s-divergence $\tilde{\mu}_N(s)$ is positive for all $s \in [0,1]$.*

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

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

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

$$
\mu_N(s) = \frac{1-s}{2} \log \det \left[\boldsymbol{\Sigma}(\delta\text{SNR}) \right] - \frac{1}{2} \log \det \left[\boldsymbol{\Sigma}(\delta\text{SNR} \times (1-s)) \right]
$$

Lemma 2. *The s-divergence in the small deviation regime can be approximated according to*

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

where the Fisher information [3] is given by

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

Proof. See Appendix B. □

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

Lemma 3. *In case of large SNR, we have*

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

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

Proof. See Appendix C. □

4. Computational Information Geometry for Classification

4.1. Formulation of the Observation Vector as a Structured Linear Model

The measurement tensor follows a noisy Q-order tensor of size $N_1 \times \ldots \times N_Q$ can be expressed as

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

where $\mathcal{N}$ is the noise tensor whose entries are assumed to be centered i.i.d. Gaussian, i.e., $[\mathcal{N}]_{n_1,\ldots,n_Q} \sim \mathcal{N}(0, \sigma^2)$ and the core tensor $\mathcal{X}$ follows either CPD or TKD given by Section 2.1.2 and Section 2.1.3, respectively. The vectorization of Equation (10) is given by

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

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

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

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

where $\mathbf{\Phi}^{\odot} = \mathbf{\Phi}^{(Q)}\odot\ldots\odot\mathbf{\Phi}^{(1)}$ is a $N\times R$ structured matrix and $\mathbf{s} = \begin{bmatrix} s_1 & \ldots & s_R \end{bmatrix}^T$ where $s_r \sim \mathcal{N}(0,\sigma_s^2)$, i.i.d. and $N = N_1\cdots N_Q$.

2. When tensor $\mathcal{X}$ follows a Q-order TKD of multilinear rank of $\{M_1,\ldots,M_Q\}$, we have

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

where $\mathbf{\Phi}^{\otimes} = \mathbf{\Phi}^{(Q)}\otimes\ldots\otimes\mathbf{\Phi}^{(1)}$ is a $N\times M$ structured matrix with $M = M_1\cdots M_Q$ and $\text{vec}(\mathcal{S})$ is the vectorization of tensor $\mathcal{S}$ 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 $\mathbf{\Phi}^{\odot} = \mathbf{\Phi}^{(Q)}\odot\ldots\odot\mathbf{\Phi}^{(1)}$ and $(\mathbf{\Phi}^{(q)})_{q=1,\ldots,Q}$ are matrices of size $N_q\times R$. In the following, we assume that matrices $\mathbf{\Phi}^{(q)}_{q=1,\ldots,Q}$ 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 1. *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*

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

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

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

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 D. ☐

Remark 2. *In [49], 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 2. *In the small* SNR *scenario, the Fisher information for CPD is given as*

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

Proof. Using Lemma 2, we can notice that

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

and that

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

converges a.s towards the second moment of the Marchenko-Pastur distribution which is $1 + c$ (see for instance [43]). □

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

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

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 3, we get immediately Equation (14). □

Remark 3. *It is interesting to note that for $c \to 0$ or 1, the optimal s-value follows the same approximated relation given by*

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

as long as 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 Equation (14) and condition SNR $\gg \exp[1]$, the desired result is proved. □

4.2.3. Approximated Analytical Expressions for $c \ll 1$ and Any SNR

In the case of low rank CPD where its rank R is supposed to be small compared to N, it is realistic to assume $c \ll 1$ since $R \ll N$.

Result 4. *Under this regime, the error exponent can be approximated as follows:*

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

Proof. See Appendix E. □

It is easy to notice that the second-order derivative of $\mu(s)$ is strictly positive. Therefore, $\mu(s)$ is a strictly convex function over interval $(0, 1)$. As a consequence, $\mu(s)$ admits at most one global

minimum. We denote by $s^\star$, the global minimizer and obtained by zeroing the first-order derivative of the error exponent. This optimal value is expressed as

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

The two following scenarios can be considered:

- At low SNR, we denote by $\mu(s^\star)$, the error exponent associated with the tightest CUB, coincides with the error exponent associated with the BUB. To see this, when $c \ll 1$, we derive the second-order approximation of the optimal value $s^\star$ in Equation (15)

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

  Result 1 and the above approximation allow us to get the best error exponent at low SNR and $c \ll 1$,

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

- Contrarily, when $\mathrm{SNR} \to \infty$, $s^\star \to 1$. As a consequence, the optimal error exponent in this regime is not the BUB anymore. Assuming that $\frac{\log \mathrm{SNR}}{\mathrm{SNR}} \to 0$, Equation (15) in Result 4 provides the following approximation of the optimal error exponent for large SNR

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

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 5. *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*

$$\frac{\mu_N(s)}{N} \overset{\mathrm{a.s}}{\longrightarrow} \mu(s) = c_1 \cdots c_Q \left[\frac{1-s}{2} \int_0^{+\infty} \cdots \int_0^{+\infty} \log(1 + \mathrm{SNR} \times \lambda_1 \cdots \lambda_Q) dv_{c_1}(\lambda_1) \cdots dv_{c_Q}(\lambda_Q) \right.$$
$$\left. - \frac{1}{2} \int_0^{+\infty} \cdots \int_0^{+\infty} \log(1 + (1-s)\mathrm{SNR} \times \lambda_1 \cdots \lambda_Q) dv_{c_1}(\lambda_1) \cdots dv_{c_Q}(\lambda_Q) \right] \tag{16}$$

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

Proof. See Appendix F. □

Remark 4. *We can notice that for $Q = 1$, the result 5 is similar to result 1. However, when $Q \geq 2$, the integrals in Equation (16) are not tractable in a closed-form expression. For instance, let $Q = 2$, we consider the integral*

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \log(1 + \text{SNR} \times \lambda_1 \lambda_2) \nu_{c_1}(d\lambda_1) \nu_{c_2}(d\lambda_2)$$

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

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., [51]). As a consequence, it cannot be expressed in closed-form. However, numerical computations can be exploited to solve efficiently the minimization problem of Equation (7).

4.3.1. Large SNR Deviation Scenario

Result 6. *In case of large* SNR, *the minimizer of Chernoff Information for TKD is given by*

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

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 3, we get immediately Equation (17). □

4.3.2. Small SNR Deviation Scenario

Under this regime, we have the following results

Result 7. *For small* SNR *deviation, the Chernoff information for the TKD is given by*

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

Proof. Using Lemma 2, we can notice that

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

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 5. *Contrary to the Remark 3, 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*

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

which does not depend on Q, and

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

which depends on Q.

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

5. Numerical Illustrations

In this section, 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.
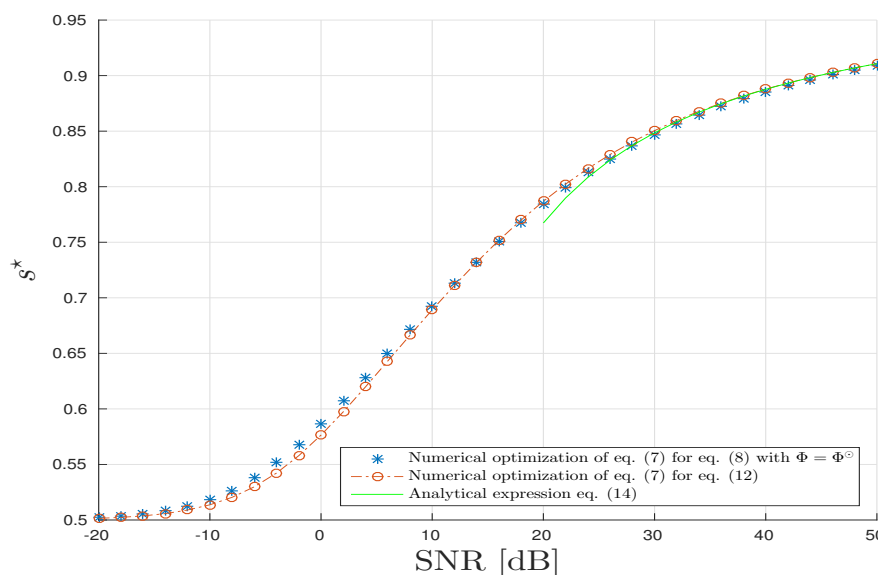
Figure 4. Canonical Polyadic Decomposition (CPD) scenario: Optimal s-parameter versus Signal to Noise Ratio (SNR) in dB.

Firstly, for the CPD model, in Figure 4, parameter $s^\star$ is drawn 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 Equation (8) with $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\odot}$. This approach has a very high computational cost especially in our asymptotic regime (for a standard computer with Intel Xeon E5-2630 2.3 GHz and 32 GB RAM, it requires 183 h to establish 10,000 simulations). The second approach is based on the numerical optimization of the closed-form expression of $\mu(s)$ given in Result 4. In this scenario, the drawback in terms of the computational cost is largely mitigated since it consists of a minimization of a 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 Equation (15). We can check that 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 $L = 10{,}000$ 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 $\hat{s}^{\star}$ is a consistent estimator of $s^{\star}$.

In Figure 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.

In Figure 6, we fix SNR $= 45$ dB and draw $s^{\star}$ obtained by Equation (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 Equation (15). The two curves approach each other as c goes to zero as predicted by our theoretical analysis.

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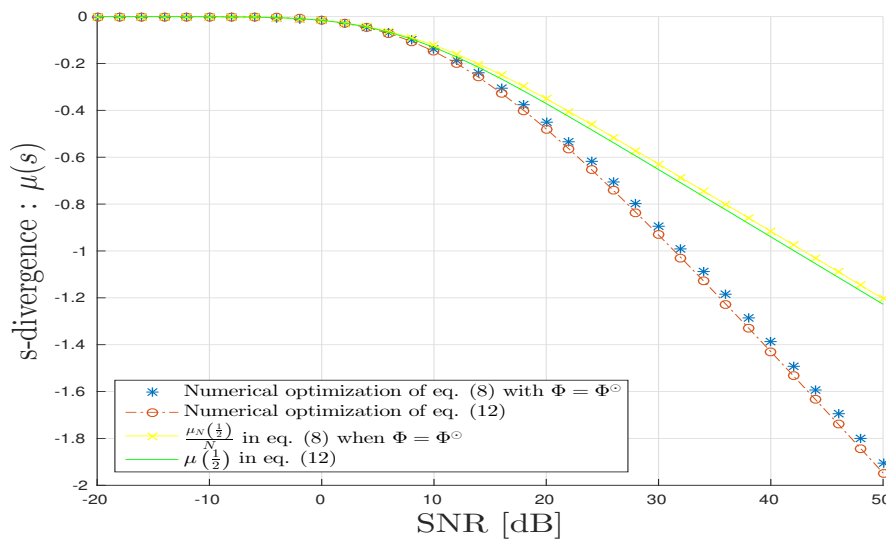

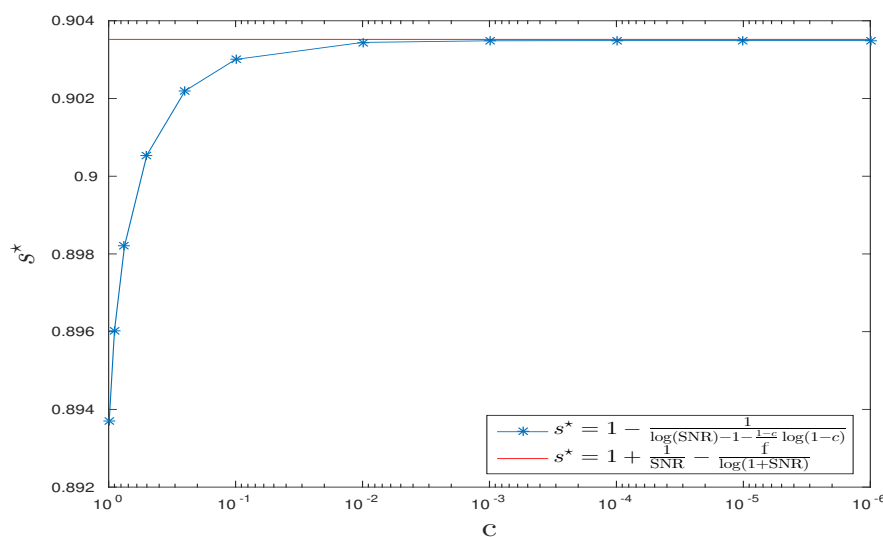
Figure 5. CPD scenario: s-divergence vs. SNR in dB.

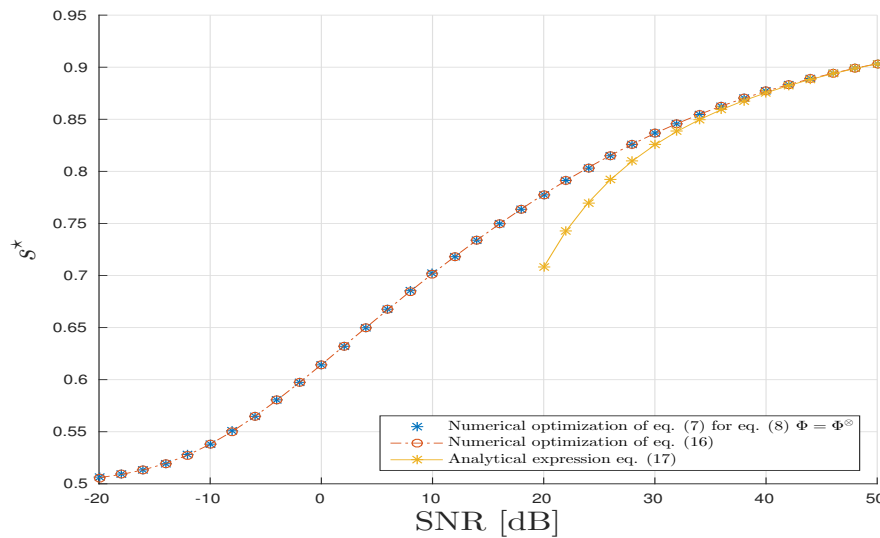**Figure 6.** CPD scenario: $s^{\star}$ vs. c, SNR $= 45$ dB.

Figure 7. TucKer Decomposition (TKD) scenario: Optimal *s*-parameter vs. SNR in dB.
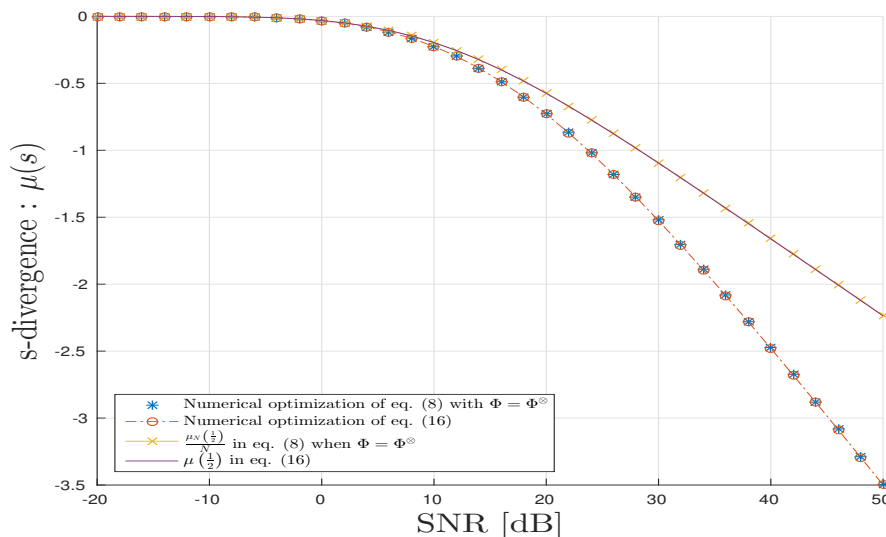

Figure 8. TKD scenario: *s*-divergence vs. SNR in dB.

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 $\boldsymbol{\Phi}^\otimes$ is $200, 200, 200 \times 100, 120, 140$ ($N = 200^3$) which is much larger than the dimension 6000×3000 of $\boldsymbol{\Phi}^\odot$ ($N = 6000$).

6. Conclusions

In this work, we derived and studied the limit performance in terms of minimal Bayes' error probability for the binary classification 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 $\hat{s}$.

Author Contributions: Gia-Thuy Pham, Rémy Boyer and Frank Nielsen contributed to the research results presented in this paper. Gia-Thuy Pham and Rémy Boyer performed the numerical experiments. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Lemma 1

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

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

is given by [15]:

$$\tilde{\mu}_N(s) = \frac{1}{2} \log \frac{\det(s\mathbf{\Sigma}_0 + (1-s)\mathbf{\Sigma}_1)}{[\det\mathbf{\Sigma}_0]^s [\det\mathbf{\Sigma}_1]^{1-s}}. \tag{A1}$$

Using the expressions of the covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, the numerator in Equation (A1) is given by

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

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

$$\log[\det \mathbf{\Sigma}_1]^{1-s} = (1-s)\left(N \log \sigma^2 + \log \det\left(\text{SNR} \times \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right)\right).$$

Using the above expressions, $\tilde{mu}_N(s)$ is given by Equation (8).

Appendix B. Proof of Lemma 2

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

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

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

$$\mu_N(s) = \frac{1-s}{2} \log \det\left[\mathbf{I} + (\delta\text{SNR}) \times \mathbf{\Phi}\mathbf{\Phi}^T\right] - \frac{1}{2} \log \det\left[\mathbf{I} + \delta\text{SNR} \times (1-s) \times \mathbf{\Phi}\mathbf{\Phi}^T\right]$$

Now, using Equation (8), and the following approximation:

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

we obtain

$$\frac{\mu_N(s)}{N} \approx (s-1)s \times \frac{(\delta\text{SNR})^2}{2} \times \frac{J_F(0)}{N}$$

where the Fisher information for $\mathbf{y}|\delta\mathrm{SNR} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\delta\mathrm{SNR}))$ is given by [3]:

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

Appendix C. Proof of Theorem 3

The first step of the proof is based on the derivation of an alternative expression of $\mu_s(\mathrm{SNR})$ given by Equation (A1) involving the inverse of the covariance matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$. Specifically, we have

$$
\begin{aligned}
\mu_s(\mathrm{SNR}) &= \frac{1}{2}\log\frac{(\det\boldsymbol{\Sigma}_0)(\det\boldsymbol{\Sigma}_1)\det((1-s)\boldsymbol{\Sigma}_0^{-1} + s\boldsymbol{\Sigma}_1^{-1})}{[\det\boldsymbol{\Sigma}_0]^s[\det\boldsymbol{\Sigma}_1]^{1-s}}\\
&= -\frac{1}{2}\log\frac{\det\left([(1-s)\boldsymbol{\Sigma}_0^{-1} + s\boldsymbol{\Sigma}_1^{-1}]^{-1}\right)}{[\det\boldsymbol{\Sigma}_0]^{1-s}[\det\boldsymbol{\Sigma}_1]^s}.
\end{aligned}
\tag{A2}
$$

The second step is to derive a closed-form expression in the high SNR regime using the following the approximation (see [52] for instance): $\left(x \times \boldsymbol{\Phi}\boldsymbol{\Phi}^T + \mathbf{I}\right)^{-1} \overset{x \gg 1}{\approx} \boldsymbol{\Pi}_{\boldsymbol{\Phi}}^{\perp} = \mathbf{I}_N - \boldsymbol{\Phi}\boldsymbol{\Phi}^{\dagger}$ where $\boldsymbol{\Pi}_{\boldsymbol{\Phi}}^{\perp}$ is an orthogonal projector such as $\boldsymbol{\Pi}_{\boldsymbol{\Phi}}^{\perp}\boldsymbol{\Phi} = \mathbf{0}$ and $\boldsymbol{\Phi}^{\dagger} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T$. The numerator in Equation (A2) is given by

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

As $s\boldsymbol{\Phi}\boldsymbol{\Phi}^{\dagger}$ 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 $s\boldsymbol{\Phi}\boldsymbol{\Phi}^{\dagger}$ can be diagonalized in the same orthonormal basis matrix, the n-th eigenvalue of the inverse of matrix $\mathbf{I}_N - s\boldsymbol{\Phi}\boldsymbol{\Phi}^{\dagger}$ is given by

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

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

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

In addition, we have

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

Finally, thanks to Equation (A2), we have

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

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

Appendix D. Proof of Result 1

The asymptotic behavior of $\frac{\mu_N(s)}{N}$ when $N_q \to +\infty$ for each $q = 1, \dots, Q$, $R \to +\infty$ in such a way that $\frac{R^{1/q}}{N_q}$ converge towards a non zero constant for each $q = 1, \dots, Q$ can be obtained thanks to large random matrix theory. We suppose that $N_1, \dots, N_Q$ 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 $c > 0$. Under this regime, the empirical eigenvalue distribution of covariance matrix $\mathbf{\Phi}^{\odot}(\mathbf{\Phi}^{\odot})^T$ is known to converge towards the so-called Marcenko–Pastur distribution. By Section 2.2, we recall that the Marcenko–Pastur distribution $\nu_c(d\lambda)$ is defined as

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

where $\lambda_c^- = (1 - \sqrt{c})^2$ and $\lambda_c^+ = (1 + \sqrt{c})^2$. We define $t_c(z) = \int_{\mathbb{R}^+} \frac{\nu_c(d\lambda)}{\lambda - z}$ the Stieltjes transform of ν_c. We have that $t_c(z)$ satisfies 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

$$t_c(-\rho) = \frac{2}{\rho - (1 - c) + \sqrt{(\rho + \lambda_c^-)(\rho + \lambda_c^+)}} \tag{A3}$$

It was established for the first time in [45] 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 $(\lambda_k)_{k=1,\dots,K}$ represent the eigenvalues of $\mathbf{X}\mathbf{X}^T$ arranged in decreasing order, then $\frac{1}{K} \sum_{k=1}^{K} \delta(\lambda - \lambda_k)$, the empirical eigenvalue distribution of $\mathbf{X}\mathbf{X}^T$ converges weakly almost surely towards ν_c, under the regime $K \to +\infty$, $P \to +\infty$, $\frac{P}{K} \to c$. In addition, we have the following property, for each continuous function $f(\lambda)$

$$\frac{1}{K} \sum_{k=1}^{K} f(\lambda_k) \xrightarrow{\text{a.s}} \int_{\mathbb{R}^+} f(\lambda) \, \nu_c(d\lambda). \tag{A4}$$

Practically, when K and P are large enough, the histogram of the eigenvalues of each realization of $\mathbf{X}\mathbf{X}^T$ accumulates around the graph of the probability density of ν_c.

The columns $(\boldsymbol{\phi}_r)_{r=1,\dots,R}$ of $\mathbf{\Phi}^{\odot}$ are vectors $(\boldsymbol{\phi}_r^{(Q)} \otimes \dots \otimes \boldsymbol{\phi}_r^{(1)})_{r=1,\dots,R}$, which are mutually independent, identically distributed, and satisfy $\mathbb{E}(\boldsymbol{\phi}_r \boldsymbol{\phi}_r^T) = \frac{I_N}{N}$. However, since the components of each column $\boldsymbol{\phi}_r$ are not independent, it results in that the entries of $\mathbf{\Phi}^{\odot}$ are not mutually independent. Applying the results of [53] (see also [54]), we can establish that the empirical eigenvalue distribution of $\mathbf{\Phi}^{\odot}(\mathbf{\Phi}^{\odot})^T$ still converges almost surely towards ν_c, under the asymptotic regime $\frac{R}{N} \to c$. For continuous function $f(\lambda) = \log(1 + \lambda/\rho)$, we apply Equation (A4), $\int_{\mathbb{R}^+} \log(1 + \lambda/\rho) \, \nu_c(d\lambda)$ can be expressed in terms of $t_c(-\rho)$ given by Equation (A3) (see e.g., [50]), we finish the proof.

Appendix E. Proof of Result 4

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, Equation (13) is

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

Using the above approximation and Equation (12), we obtain Result 4.

Appendix F. Proof of Result 5

We first denote $\lambda_1^{(q)} \geq \lambda_2^{(q)} \geq ... \geq \lambda_{n_q}^{(q)} \geq ... \geq \lambda_{N_q}^{(q)}$ 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 $M = M_1 ... M_Q$, $N = N_1 ... N_Q$, we obtain immediately that

$$\frac{1}{N} \log \det \left(\text{SNR} \times \boldsymbol{\Phi}^{\otimes}(\boldsymbol{\Phi}^{\otimes})^T + \mathbf{I} \right) = \frac{1}{N} \sum_{n_1=1}^{N_1} ... \sum_{n_q=1}^{N_Q} \log \left(\text{SNR} \times \lambda_{n_1}^{(1)} \cdots \lambda_{n_Q}^{(Q)} + 1 \right)$$

$$= \frac{M}{N} \frac{1}{M} \sum_{n_1=1}^{M_1} ... \sum_{n_q=1}^{M_Q} \log \left(\text{SNR} \times \lambda_{n_1}^{(1)} \cdots \lambda_{n_Q}^{(Q)} + 1 \right)$$

and that

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

Similarly, we have that

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

We obtain easily Result 5.

References

1. Besson, O.; Scharf, L.L. CFAR matched direction detector. *IEEE Trans. Signal Process.* **2006**, *54*, 2840–2844.
2. Bianchi, P.; Debbah, M.; Maida, M.; Najim, J. Performance of Statistical Tests for Source Detection using Random Matrix Theory. *IEEE Trans. Inf. Theory* **2011**, *57*, 2400–2419.
3. Kay, S.M. *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*; PTR Prentice-Hall: Englewood Cliffs, NJ, USA, 1993.
4. Loubaton, P.; Vallet, P. Almost Sure Localization of the Eigenvalues in a Gaussian Information Plus Noise Model. Application to the Spiked Models. *Electron. J. Probab.* **2011**, *16*, 1934–1959.
5. Mestre, X. Improved Estimation of Eigenvalues and Eigenvectors of Covariance Matrices Using Their Sample Estimates. *IEEE Trans. Inf. Theory* **2008**, *54*, 5113–5129.
6. Baik, J.; Silverstein, J. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivar. Anal.* **2006**, *97*, 1382–1408.
7. Silverstein, J.W.; Combettes, P.L. Signal detection via spectral theory of large dimensional random matrices. *IEEE Trans. Signal Process.* **1992**, *40*, 2100–2105.
8. Cheng, Y.; Hua, X.; Wang, H.; Qin, Y.; Li, X. The Geometry of Signal Detection with Applications to Radar Signal Processing. *Entropy* **2016**, *18*, 381.

9. Ali, S.M.; Silvey, S.D. A General Class of Coefficients of Divergence of One Distribution from Another. *J. R. Stat. Soc. Ser. B (Methodol.)* **1966**, *28*, 131–142.

10. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.

11. Kailath, T. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Trans. Commun. Technol.* **1967**, *15*, 52–60.

12. Nielsen, F. *Hypothesis Testing, Information Divergence and Computational Geometry*; Geometric Science of Information; Springer: Berlin, Germany, 2013; pp. 241–248.

13. Sinanovic, S.; Johnson, D.H. Toward a theory of information processing. *Signal Process.* **2007**, *87*, 1326–1344.

14. Chernoff, H. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations. *Ann. Math. Stat.* **1952**, *23*, 493–507.

15. Nielsen, F. Chernoff information of exponential families. *arXiv* **2011**, arXiv:1102.2684.

16. Chepuri, S.P.; Leus, G. Sparse sensing for distributed Gaussian detection. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015.

17. Tang, G.; Nehorai, A. Performance Analysis for Sparse Support Recovery. *IEEE Trans. Inf. Theory* **2010**, *56*, 1383–1399.

18. Lee, Y.; Sung, Y. Generalized Chernoff Information for Mismatched Bayesian Detection and Its Application to Energy Detection. *IEEE Signal Process. Lett.* **2012**, *19*, 753–756.

19. Grossi, E.; Lops, M. Space-time code design for MIMO detection based on Kullback-Leibler divergence. *IEEE Trans. Inf. Theory* **2012**, *58*, 3989–4004.

20. Sen, S.; Nehorai, A. Sparsity-Based Multi-Target Tracking Using OFDM Radar. *IEEE Trans. Signal Process.* **2011**, *59*, 1902–1906.

21. Boyer, R.; Delpha, C. Relative-entropy based beamforming for secret key transmission. In Proceedings of the 2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM), Hoboken, NJ, USA, 17–20 June 2012.

22. Tran, N.D.; Boyer, R.; Marcos, S.; Larzabal, P. Angular resolution limit for array processing: Estimation and information theory approaches. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012.

23. Katz, G.; Piantanida, P.; Couillet, R.; Debbah, M. Joint estimation and detection against independence. In Proceedings of the Annual Conference on Communication Control and Computing (Allerton), Monticello, IL, USA, 30 September–3 October 2014; pp. 1220–1227.

24. Nielsen, F. An information-geometric characterization of Chernoff information. *IEEE Signal Process. Lett.* **2013**, *20*, 269–272.

25. Cichocki, A.; Mandic, D.; De Lathauwer, L.; Zhou, G.; Zhao, Q.; Caiafa, C.; Phan, H.A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Process. Mag.* **2015**, *32*, 145–163.

26. Comon, P. Tensors: A brief introduction. *IEEE Signal Process. Mag.* **2014**, *31*, 44–53.

27. De Lathauwer, L.; Moor, B.D.; Vandewalle, J. A Multilinear Singular Value Decomposition. *SIAM J. Matrix Anal. Appl.* **2000**, *21*, 1253–1278.

28. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311.

29. Comon, P.; Berge, J.T.; De Lathauwer, L.; Castaing, J. Generic and Typical Ranks of Multi-Way Arrays. *Linear Algebra Appl.* **2009**, *430*, 2997–3007.

30. De Lathauwer, L. A survey of tensor methods. In Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS 2009, Taipei, Taiwan, 24–27 May 2009.

31. Comon, P.; Luciani, X.; De Almeida, A.L.F. Tensor decompositions, alternating least squares and other tales. *J. Chemom.* **2009**, *23*, 393–405.

32. Goulart, J.H.D.M.; Boizard, M.; Boyer, R.; Favier, G.; Comon, P. Tensor CP Decomposition with Structured Factor Matrices: Algorithms and Performance. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 757–769.

33. Eckart, C.; Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1936**, *1*, 211–218.

34. Badeau, R.; Richard, G.; David, B. Fast and stable YAST algorithm for principal and minor subspace tracking. *IEEE Trans. Signal Process.* **2008**, *56*, 3437–3446.

35. Boyer, R.; Badeau, R. Adaptive multilinear SVD for structured tensors . In Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'06), Toulouse, France, 14–19 May 2006.

36. Boizard, M.; Ginolhac, G.; Pascal, F.; Forster, P. Low-rank filter and detector for multidimensional data based on an alternative unfolding HOSVD: Application to polarimetric STAP. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 119.

37. Bouleux, G.; Boyer, R. Sparse-Based Estimation Performance for Partially Known Overcomplete Large-Systems. *Signal Process.* **2017**, *139*, 70–74.

38. Boyer, R.; Couillet, R.; Fleury, B.-H.; Larzabal, P. Large-System Estimation Performance in Noisy Compressed Sensing with Random Support—A Bayesian Analysis. *IEEE Trans. Signal Process.* **2016**, *64*, 5525–5535.

39. Ollier, V.; Boyer, R.; El Korso, M.N.; Larzabal, P. Bayesian Lower Bounds for Dense or Sparse (Outlier) Noise in the RMT Framework. In Proceedings of the 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 16), Rio de Janerio, Brazil, 10–13 July 2016.

40. Wishart, J. The generalized product moment distribution in samples. *Biometrika* **1928**, *20A*, 32–52.

41. Wigner, E.P. On the statistical distribution of the widths and spacings of nuclear resonance levels. *Proc. Camb. Philos. Soc.* **1951**, *47*, 790–798.

42. Wigner, E.P. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **1955**, *62*, 548–564.

43. Bai, Z.D.; Silverstein, J.W. *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed.; Springer Series in Statistics; Springer: Berlin, Germany, 2010.

44. Girko, V.L. *Theory of Random Determinants*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990.

45. Marchenko, V.A.; Pastur, L.A. Distribution of eigenvalues for some sets of random matrices. *Math. Sb. (N.S.)* **1967**, *72*, 507–536.

46. Voiculescu, D. Limit laws for random matrices and free products. *Invent. Math.* **1991**, *104*, 201–220.

47. Boyer, R.; Nielsen, F. Information Geometry Metric for Random Signal Detection in Large Random Sensing Systems. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.

48. Boyer, R.; Loubaton, P. Large deviation analysis of the CPD detection problem based on random tensor theory. In Proceedings of the 2017 25th European Association for Signal Processing (EUSIPCO), Kos, Greece, 28 August–2 September 2017.

49. Lytova, A. Central Limit Theorem for Linear Eigenvalue Statistics for a Tensor Product Version of Sample Covariance Matrices. *J. Theor. Prob.* **2017**, 1–34.

50. Tulino, A.M.; Verdu, S. *Random Matrix Theory and Wireless Communications*; Now Publishers Inc.: Hanover, MA, USA, 2004; Volume 1.

51. Milne-Thomson, L.M. "Elliptic Integrals" (Chapter 17). In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing; Abramowitz, M., Stegun, I.A., Eds.; Dover Publications: New York, NY, USA, 1972; pp. 587–607.

52. Behrens, R.T.; Scharf, L.L. Signal processing applications of oblique projection operators. *IEEE Trans. Signal Process.* **1994**, *42*, 1413–1424.

53. Pajor, A.; Pastur, L.A. On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution. *Stud. Math.* **2009**, *195*, 11–29.

54. Ambainis, A.; Harrow, A.W.; Hastings, M.B. Random matrix theory: Extending random matrix theory to mixtures of random product states. *Commun. Math. Phys.* **2012**, *310*, 25–74.