On the Error Exponent of a Random Tensor with Orthonormal Factor Matrices

Rémy Boyer[1] and Frank Nielsen[2] *

[1] University of Paris-Sud, L2S, Department of Signals and Statistics, France
remy.boyer@l2s.centralesupelec.fr
[2] École Polytechnique, LIX, France
Frank.Nielsen@acm.org

Abstract. In signal processing, the detection error probability of a random quantity is a fundamental and often difficult problem. In this work, we assume that we observe under the alternative hypothesis a noisy tensor admitting a Tucker decomposition parametrized by a set of orthonormal factor matrices and a random core tensor of interest with fixed multilinear ranks. The detection of the random entries of the core tensor is hard to study since an analytic expression of the error probability is not tractable. To cope with this difficulty, the Chernoff Upper Bound (CUB) on the error probability is studied for this tensor-based detection problem. The tightest CUB is obtained for the minimal error exponent value, denoted by $s^\star$, that requires a costly numerical optimization algorithm. An alternative strategy to upper bound the error probability is to consider the Bhattacharyya Upper Bound (BUB) by prescribing $s^\star = 1/2$. In this case, the costly numerical optimization step is avoided but no guarantee exists on the tightness optimality of the BUB. In this work, a simple analytical expression of $s^\star$ is provided with respect to the Signal to Noise Ratio (SNR). Associated to a compact expression of the CUB, an easily tractable expression of the tightest CUB is provided and studied. A main conclusion of this work is that the BUB is the tightest bound at low SNRs but this property is no longer true at higher SNRs.

1 Introduction

The theory of tensor decomposition is an important research topic (see for instance [1, 2]). They are useful to extract relevant information confined into a small dimensional subspaces from a massive volume of measurements while reducing the computational cost. In the context where the measurements are naturally modeled according to more than two axes of variations, *i.e.*, in the case of tensors, the problem of obtaining a *low rank approximation* faces a number of practical and fundamental difficulties. Indeed, even if some aspects of the tensor algebra can be considered as mature, several "obvious" algebraic concepts in the

* This research was partially supported by Labex DigiCosme (project ANR-11-LABEX-0045-DIGICOSME) operated by ANR as part of the program "Investissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02).

matrix case such as decomposition uniqueness, rank, or the notions of singular and eigen-values remain active and challenging research areas [3]. The Tucker decomposition [4] and the HOSVD (High-Order SVD) [5] are two popular decompositions being an alternative to the Canonical Polyadic decomposition [6]. In this case, the notion of tensorial rank is no longer relevant and an alternative rank definition is used. Specifically, it is standard to use the multilinear ranks defined as the set of strictly positive integers $\{R_1, R_2, R_3\}$ where R_p is the usual rank (in the matrix sense) of the p-th mode or unfolding matrix. 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 [7] or adaptively [8]. Unfortunately, it is shown that the fixed (multilinear) rank tensor based on this procedure is generally suboptimal in the Fröbenius norm sense [5]. In other words, there does not exist a generalization of the Eckart-Young theorem for tensor of order strictly greater than two. Despite of this theoretical singularity, we focus our effort in the detection performance of a given multilinear rank tensor following the Tucker model with orthonormal factor matrices, leading to the HOSVD. It is important to note that the detection theory for tensors is an under-studied research topic. To the best of our knowledge, only the publication [9] tackles this problem in the context of RADAR multidimensional data detection. A major difference with this publication is that their analysis is dedicated to the performance of a low rank detection after matched filtering. More specifically, the goal is to decompose a 3-order tensor $\boldsymbol{\mathcal{X}}$ of size $N_1 \times N_2 \times N_3$ into a core tensor denoted by $\boldsymbol{\mathcal{S}}$ of size $R_1 \times R_2 \times R_3$ and into three rank-R_p orthonormal factor matrices $\{\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \boldsymbol{\Phi}_3\}$ each of size $N_p \times R_p$ with $R_p < N_p$, $\forall p$. 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. The *binary hypothesis test* can be described under the null hypothesis $\mathcal{H}_0 : \text{SNR} = 0$ (*i.e.*, only the noise is present) and the alternative hypothesis $\mathcal{H}_1 : \text{SNR} \neq 0$ (*i.e.*, there exists a signal of interest). First note that the exact derivation of the analytical expression of the error probability is not tractable in an analytical way even in the matrix case [10]. To tackle this problem, we adopt an information-geometric characterization of the detection performance [11, 12].

2 Chernoff information framework

2.1 The Bayes' detection theory

Let $\Pr(\mathcal{H}_i)$ be the *a priori* hypothesis probability with $\Pr(\mathcal{H}_0) + \Pr(\mathcal{H}_1) = 1$. Let $p_i(\mathbf{y}) = p(\mathbf{y}|\mathcal{H}_i)$ and $\Pr(\mathcal{H}_i|\mathbf{y})$ be the i-th *conditional* and the *posterior* probabilities, respectively. The Bayes' detection rule chooses the hypothesis $\mathcal{H}_i$ associated with the largest posterior probability $\Pr(\mathcal{H}_i|\mathbf{y})$. Introduce the indicator hypothesis function according to

$$\phi(\mathbf{y}) \sim \text{Bernou}(\alpha),$$

where Bernou(α) stands for the Bernoulli distribution of success probability $\alpha = \Pr(\phi(\mathbf{y}) = 1) = \Pr(\mathcal{H}_1)$. Function $\phi(\mathbf{y})$ is defined on $\mathcal{X} \to \{0, 1\}$ where $\mathcal{X}$ is the data-set enjoying the following decomposition $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$ where $\mathcal{X}_0 = \{\mathbf{y} : \phi(\mathbf{y}) = 0\} = \mathcal{X} \setminus \mathcal{X}_1$ and

$$
\begin{aligned}
\mathcal{X}_1 &= \{\mathbf{y} : \phi(\mathbf{y}) = 1\} \\
&= \left\{\mathbf{y} : \Omega(\mathbf{y}) = \log \frac{\Pr(\mathcal{H}_1|\mathbf{y})}{\Pr(\mathcal{H}_0|\mathbf{y})} > 0\right\} \\
&= \left\{\mathbf{y} : \Lambda(\mathbf{y}) = \log \frac{p_1(\mathbf{y})}{p_0(\mathbf{y})} > \log \tau\right\}
\end{aligned}
$$

in which $\tau = \frac{1-\alpha}{\alpha}$, $\Omega(\mathbf{y})$ is the log posterior-odds ratio and $\Lambda(\mathbf{y})$ is the log-likelihood ratio. The average error probability is defined as

$$
P_e^{(N)} = \mathbb{E}\{\Pr(\text{Error}|\mathbf{y})\}, \tag{1}
$$

with

$$
\Pr(\text{Error}|\mathbf{y}) = \begin{cases} \Pr(\mathcal{H}_0|\mathbf{y}) \; if \; \mathbf{y} \in \mathcal{X}_1, \\ \Pr(\mathcal{H}_1|\mathbf{y}) \; if \; \mathbf{y} \in \mathcal{X}_0. \end{cases}
$$

2.2 Chernoff Upper Bound (CUB)

Using the fact that $\min\{a, b\} \leq a^s b^{1-s}$ with $s \in (0, 1)$ and $a, b > 0$ in eq. (1), the minimal error probability is upper bounded as follows

$$
P_e^{(N)} \leq \frac{\alpha}{\tau^s} \cdot \int_{\mathcal{X}} p_0(\mathbf{y})^{1-s} p_1(\mathbf{y})^s d\mathbf{y} = \frac{\alpha}{\tau^s} \cdot \exp[-\mu_s], \tag{2}
$$

where the Chernoff s-divergence is defined according to

$$
\mu_s = -\log M_{\Lambda(\mathbf{y}|\mathcal{H}_0)}(s),
$$

with $M_{\Lambda(\mathbf{y}|\mathcal{H}_0)}(s)$ the M oment Generating Function (mgf) of the log-likelihood ratio under the null hypothesis.

The Chernoff Upper Bound (CUB) of the error probability is given by

$$
P_e^{(N)} \leq \frac{\alpha}{\tau^s} \cdot \exp[-\mu_s] = \alpha \cdot \exp[-r_s] \tag{3}
$$

where

$$
r_s = \mu_s + s \log \tau
$$

is the exponential decay rate of the CUB. Assume that an optimal value of s denoted by $s^\star \in (0, 1)$ exists then the tightest CUB verifies

$$
P_e^{(N)} \leq \alpha \cdot \exp[-r_{s^\star}] < \alpha \cdot \exp[-r_s].
$$

The above condition is equivalent to maximize the exponential decay rate, *i.e.*,

$$s^\star = \arg\max_{s\in(0,1)} r_s. \tag{4}$$

Finally using eq. (3) and eq. (4), we obtain the Chernoff Upper Bound. The Bhattacharyya Upper Bound (BUB) is obtained by eq. (3) by fixing $s = 1/2$ instead of solving eq. (4). The two bounds verify the following inequality relation:

$$P_e^{(N)} \le \alpha \cdot \exp[-r_{s^\star}] \le \alpha \cdot \exp\left[-r_{\frac{1}{2}}\right].$$

Note that in many encountered problems, the two hypothesis are assumed to be equi-probable, *i.e.*, $\alpha = 1/2$ and $\tau = 1$. Then the exponential decay rate is in this scenario given by the Chernoff s-divergence since $r_s = \mu_s$ and the tightest CUB is

$$P_e^{(N)} \le \frac{1}{2}\exp[-\mu_{s^\star}].$$

3 Tensor detection with orthonormal factors

3.1 Binary hypothesis test formulation for random tensors

Tucker model with orthonormal factors. Assume that the multidimensional measurements follow a noisy 3-order tensor of size $N_1 \times N_2 \times N_3$ given by

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

where $\boldsymbol{\mathcal{N}}$ is the $N_1 \times N_2 \times N_3$ is the noise tensor where each entry is centered *i.i.d.* Gaussian, *i.e.* $[\boldsymbol{\mathcal{N}}]_{n_1,n_2,n_3} \sim \mathcal{N}(0,\sigma^2)$ and

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{\Phi}_1 \times_2 \boldsymbol{\Phi}_2 \times_3 \boldsymbol{\Phi}_3 \tag{5}$$

is the $N_1 \times N_2 \times N_3$ "data" tensor following a Tucker model of (R_1, R_2, R_3)-multilinear rank. Matrices $\{\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \boldsymbol{\Phi}_3\}$ are the three orthonormal factors each of size $N_p \times R_p$ with $N_p > R_p$. These factors are for instance involved in the Higher-Order SVD (HOSVD) [5] with $\boldsymbol{\Phi}_p^T \boldsymbol{\Phi}_p = \mathbf{I}_{R_p}$ and $\boldsymbol{\Pi}_p = \boldsymbol{\Phi}_p \boldsymbol{\Phi}_p^T$ a $N_p \times N_p$ orthogonal projector on the range space of $\boldsymbol{\Phi}_p$. The $R_1 \times R_2 \times R_3$ core tensor is given by

$$\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{\Phi}_1^T \times_2 \boldsymbol{\Phi}_2^T \times_3 \boldsymbol{\Phi}_3^T.$$

Formulating the detection test. We assume that each entry of the core tensor is centered *i.i.d.* Gaussian, *i.e.* $[\boldsymbol{\mathcal{S}}]_{r_1,r_2,r_3} \sim \mathcal{N}(0,\sigma_{\mathbf{s}}^2)$. Let $\mathbf{Y}_n$ be the n-th frontal $N_1 \times N_2$ slab of the 3-order tensor $\boldsymbol{\mathcal{Y}}$, the vectorized tensor expression is defined according to

$$\mathbf{y} = \left[(\mathrm{vec}\mathbf{Y}_1)^T \ldots (\mathrm{vec}\mathbf{Y}_{N_3})^T\right]^T = \mathbf{x} + \mathbf{n} \in \mathbb{R}^{N\times 1}$$

where $N = N_1 \cdot N_2 \cdot N_3$, $\mathbf{n}$ is the vectorization of the noise tensor $\boldsymbol{\mathcal{N}}$ and

$$\mathbf{x} = \boldsymbol{\Phi}\mathbf{s}$$

with $\mathbf{s}$ the vectorization of the core tensor $\boldsymbol{\mathcal{S}}$ and

$$\boldsymbol{\Phi} = \boldsymbol{\Phi}_3 \otimes \boldsymbol{\Phi}_2 \otimes \boldsymbol{\Phi}_1$$

is a $N \times R$ structured matrix with $R = R_1 \cdot R_2 \cdot R_3$.

In this framework, the associated equi-probable binary hypothesis test for the detection of the random signal, $\mathbf{s}$, is

$$\begin{cases} \mathcal{H}_0 : \mathbf{y}\big|\boldsymbol{\Phi}, \sigma^2 \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_0 = \sigma^2\mathbf{I}_N\right), \\ \mathcal{H}_1 : \mathbf{y}\big|\boldsymbol{\Phi}, \sigma^2 \sim \mathcal{N}\Big(\mathbf{0}, \boldsymbol{\Sigma}_1 = \sigma^2\left(\mathrm{SNR}\cdot\boldsymbol{\Pi} + \mathbf{I}_N\right)\Big) \end{cases}$$

where $\mathrm{SNR} = \sigma_{\mathbf{s}}^2/\sigma^2$ is the signal to noise ratio and $\boldsymbol{\Pi} = \boldsymbol{\Pi}_3 \otimes \boldsymbol{\Pi}_2 \otimes \boldsymbol{\Pi}_1$ is an orthogonal projector. The performance of the detector of interest is quite often difficult to determine analytically [10]. As a consequence, we adopt the methodology of the CUB to upper bound it.

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

$$\begin{aligned}\mathbb{E}_{\mathbf{y}|\hat{\mathcal{H}}}\Lambda(\mathbf{y}) &= \int_{\mathcal{X}} p(\mathbf{y}\big|\hat{\mathcal{H}}) \log\frac{p_1(\mathbf{y})}{p_0(\mathbf{y})}d\mathbf{y} \\ &= \int_{\mathcal{X}} p(\mathbf{y}\big|\hat{\mathcal{H}}) \log\left[\frac{p(\mathbf{y}\big|\hat{\mathcal{H}})}{p_0(\mathbf{y})}\cdot\left(\frac{p(\mathbf{y}\big|\hat{\mathcal{H}})}{p_1(\mathbf{y})}\right)^{-1}\right]d\mathbf{y} \\ &= \mathcal{KL}(\hat{\mathcal{H}}, \mathcal{H}_0) - \mathcal{KL}(\hat{\mathcal{H}}, \mathcal{H}_1)\end{aligned}$$

where the Kullback-Leibler pseudo-distances are

$$\mathcal{KL}(\hat{\mathcal{H}}, \mathcal{H}_0) = \int_{\mathcal{X}} p(\mathbf{y}\big|\hat{\mathcal{H}}) \log\frac{p(\mathbf{y}\big|\hat{\mathcal{H}})}{p_0(\mathbf{y})}d\mathbf{y}, \quad \mathcal{KL}(\hat{\mathcal{H}}, \mathcal{H}_1) = \int_{\mathcal{X}} p(\mathbf{y}\big|\hat{\mathcal{H}}) \log\frac{p(\mathbf{y}\big|\hat{\mathcal{H}})}{p_1(\mathbf{y})}d\mathbf{y}.$$

The corresponding data-space for hypothesis $\mathcal{H}_1$ is

$$\mathcal{X}_1 = \{\mathbf{y} : \Lambda(\mathbf{y}) > \tau'\}$$

with

$$\Lambda(\mathbf{y}) = \mathbf{y}^T(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1})\mathbf{y} = \frac{1}{\sigma^2}\mathbf{y}^T\boldsymbol{\Phi}\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathrm{SNR}\cdot\mathbf{I}\right)^{-1}\boldsymbol{\Phi}^T\mathbf{y}$$

$$\tau' = \log\frac{\det(\boldsymbol{\Sigma}_0)}{\det(\boldsymbol{\Sigma}_1}) = -\log\det\left(\mathrm{SNR}\cdot\boldsymbol{\Pi} + \mathbf{I}_N\right)$$

where $\det(\cdot)$ stands for the determinant. Thus, the alternative hypothesis is selected, *i.e.*, $\hat{\mathcal{H}} = \mathcal{H}_1$ if

$$\mathbb{E}_{\mathbf{y}|\hat{\mathcal{H}}}\Lambda(\mathbf{y}) = \frac{1}{\sigma^2}\mathrm{Tr}\left\{\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathrm{SNR}\cdot\mathbf{I}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Sigma}\boldsymbol{\Phi}\right\} > \tau'$$

or equivalently

$$\mathcal{KL}(\hat{\mathcal{H}}, \mathcal{H}_0) > \tau' + \mathcal{KL}(\hat{\mathcal{H}}, \mathcal{H}_1).$$

4 Tightest CUB

4.1 Derivation of the bound

Theorem 1. *Let $c = R/N < 1$. The Chernoff s-divergence for the above test is given by*

$$\mu_s = \frac{c}{2}\left((1-s)\cdot\log(\mathrm{SNR}+1) - \log\left(\mathrm{SNR}\cdot(1-s)+1\right)\right).$$

Proof. According to [13], the Chernoff s-divergence for the above test is given by

$$\mu_s = \frac{1-s}{2N}\log\det\left(\mathrm{SNR}\cdot\boldsymbol{\Pi} + \mathbf{I}\right) - \frac{1}{2N}\log\det\left(\mathrm{SNR}\cdot(1-s)\boldsymbol{\Pi} + \mathbf{I}\right).$$

Using $\lambda\{\boldsymbol{\Pi}\} = \{\underbrace{1,\ldots,1}_{R}, \underbrace{0,\ldots,0}_{N-R}\}$ in the above expression yields Theorem 1.

Theorem 2. 1. *The Chernoff s-divergence is a strictly convex function and admits an unique minimizer given by*

$$s^\star = \frac{1}{\mathrm{SNR}}\left(1 + \mathrm{SNR} - \frac{1}{\psi(\mathrm{SNR})}\right) \tag{6}$$

where $\psi(\mathrm{SNR}) = \frac{\log(\mathrm{SNR}+1)}{\mathrm{SNR}}$.

2. *The tightest CUB for the (R_1, R_2, R_3)-multilinear rank orthonormal Tucker decomposition of eq.* (5) *is given by*

$$\mu_{s^\star} = \frac{c}{2}\Big(1 - \psi(\mathrm{SNR}) + \log\psi(\mathrm{SNR})\Big).$$

Proof. The proof is straightforward and thus omitted due to the lack of space.

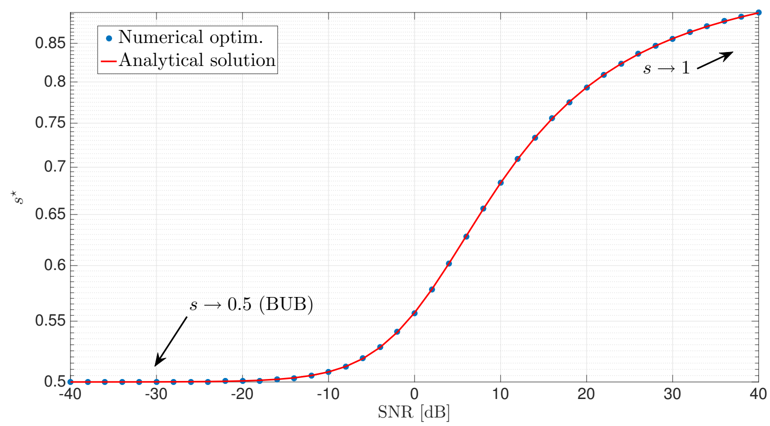

Fig. 1. Optimal s-value *vs* SNR in dB for a $(3,3,3)$-multilinear rank tensor $\boldsymbol{\mathcal{X}}$ of size $4 \times 4 \times 4$: The exact analytical formula is in full agreement with the numerical approximation scheme.

4.2 Analysis in typical limit regimes

We can identify the two following limit scenarii:

- At low SNR, the tightest divergence, denoted by $\mu_{s^\star}$, coincides with the divergence $\mu_{1/2}$ associated with the BUB. Indeed, the optimal value in (6) admits a second-order approximation for SNR $\ll 1$ according to

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

- At contrary for SNR $\gg 1$, we have $s^\star \approx 1$. So, the BUB is a loose bound in this regime.

To illustrate our analysis, on Fig. 1, the optimal s value obtained thanks to a numerical optimization of eq. (4) using the divergence given in Theorem 1 and the analytical solution reported in eq. (6) are plotted. We can check that the predicted analytical optimal s-value is in agreement with the approximated numerical one. We also verify the s-value in the the low and high SNR regimes. In particular, for high SNRs, the optimal value is far from $1/2$.

5 Conclusion

Performance detection in terms of the minimal Bayes' error probability for multi-dimensional measurements is a fundamental problem at the heart of many challenging applications. Interestingly, this tensor detection problem has received little attention so far. In this work, we derived analytically a tightest upper bound on the minimal Bayes' error probability for the detection of a random core tensor denoted by $\boldsymbol{\mathcal{S}}$ given a $N_1 \times N_2 \times N_3$ noisy observation tensor $\boldsymbol{\mathcal{X}}$ following an orthonormal Tucker model with a (R_1, R_2, R_3)-multilinear rank with $R_p < N_p$, $\forall p$. In particular, we showed that the tightest upper bound in the high SNR regime is not the Bhattacharyya upper bound.

References

1. Comon, P.: Tensors: A brief introduction. IEEE Signal Processing Magazine **31**(3) (2014) 44–53
2. 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 Processing Magazine **32**(2) (2015) 145–163
3. Chang, K., Qi, L., Zhang, T.: A survey on the spectral theory of nonnegative tensors. Numerical Linear Algebra with Applications **20**(6) (2013) 891–912
4. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3) (1966) 279–311
5. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM journal on Matrix Analysis and Applications **21**(4) (2000) 1253–1278
6. Bro, R.: Parafac: Tutorial and applications. Chemometrics and intelligent laboratory systems **38**(2) (1997) 149–171
7. Badeau, R., Boyer, R.: Fast multilinear singular value decomposition for structured tensors. SIAM Journal on Matrix Analysis and Applications **30**(3) (2008) 1008–1021
8. Boyer, R., Badeau, R.: Adaptive multilinear SVD for structured tensors. In: IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. Volume 3., IEEE (2006) III–III
9. 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 Journal on Advances in Signal Processing **2014**(1) (2014) 1–14
10. Kay, S.M.: Fundamentals of statistical signal processing: Estimation theory. PTR Prentice-Hall, Englewood Cliffs, NJ (1993)
11. Nielsen, F., Bhatia, R.: Matrix information geometry. Springer (2013)
12. Cover, T.M., Thomas, J.A.: Elements of information theory. John Wiley & Sons (2012)
13. Boyer, R., Nielsen, F.: Information geometry metric for random signal detection in large random sensing systems. ICASSP (2017)