Large Deviation Analysis of the CPD Detection Problem Based on Random Tensor Theory

Rémy Boyer

Université Paris-Sud / Laboratoire des Signaux et Systèmes

GDR ISIS
"Entropies, divergences et mesures informationnelles classiques et généralisées"

Context of this work

Collaborators

- Frank Nielsen (LIX)
- Philippe Loubaton (University of Marne La Vallée)
- Gia-Thuy Pham (L2S, post-doc)

Related publications

- Boyer, R., and Loubaton, P., "Large Deviation Analysis of the CPD Detection Problem Based on Random Tensor Theory", *Proc. of European Signal Processing Conference (EUSIPCO'17), invited article.*
- Boyer, R., and Nielsen F., "Information Geometry Metric for Random Signal Detection in Large Random Sensing Systems", *IEEE, Proc. of International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'17)*.

The classification framework

- The performance in terms of minimal Bayes' error probability for detection of a random tensor is a fundamental understudied difficult problem.
- The detection of the random entries of the core tensor is hard to study since an analytic expression of the error probability is not easily tractable.

$\rightarrow$ Chernoff Upper Bound (CUB) and the error exponent in the doubly asymptotic regime, *i.e.*, tensor sizes and number of parameter of interest grows jointly in a control way.

Noisy rank-R CPD model with random amplitudes

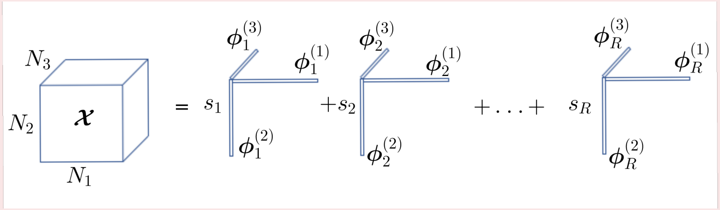
Assume that the multidimensional measurement tensor follows a noisy Q-order tensor of size $N_1 \times \ldots \times N_Q$ given by

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} \tag{1}$$

where $[\mathcal{N}]_{n_1,\ldots,n_Q} \sim \mathcal{N}(0, \sigma^2)$, *i.i.d.*, and

Random rank-R CPD of order Q

$$\mathcal{X} = \sum_{r=1}^{R} s_r \underbrace{\left(\phi_r^{(1)} \circ \ldots \circ \phi_r^{(Q)}\right)}_{\mathcal{X}_r} \quad \text{with} \quad \begin{cases} \operatorname{rank} \mathcal{X}_r = 1 \\ s_r \sim \mathcal{N}(0, \sigma_s^2), \textit{i.i.d.} \end{cases} \tag{2}$$



Random structured linear system and vectorized tensor

The vectorization of $\mathcal{Y}$ is given by

$$\mathbf{y}_N = \mathrm{vec}\boldsymbol{\mathcal{Y}} = \mathbf{x} + \mathbf{n}, \quad \text{with} \begin{cases} N = N_1 \cdots N_Q \\ \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), i.i.d. \end{cases} \tag{3}$$

where

Random rank-R CPD of order Q

$$\mathbf{x} = \mathrm{vec}\boldsymbol{\mathcal{X}} = \underbrace{\boldsymbol{\Phi}_Q \odot \ldots \odot \boldsymbol{\Phi}_1}_{\boldsymbol{\Phi} \quad (N \times R)} \underbrace{\begin{bmatrix} s_1 \\ \vdots \\ s_R \end{bmatrix}}_{\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{e}}^2 \mathbf{I}_N)} \tag{4}$$

with the q-th $N_q \times R$ factor matrix given by

$$\boldsymbol{\Phi}_q = [\phi_1^{(q)} \ldots \phi_R^{(q)}] \tag{5}$$

where $N, R \to \infty$ with a finite aspect ratio $R/N \to c \in (0, 1)$.

Bayes' theory detection

- Equi-probable Gaussian distributed hypothesis detection test :

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

- The error probability given $\mathbf{y}_N$ for the above test is

$$\Pr(\mathrm{Error}|\mathbf{y}_N) = \begin{cases} \Pr(\mathcal{H}_0|\mathbf{y}_N) & \textit{if} \quad \mathbf{y}_N \in \mathcal{X}_1, \\ \Pr(\mathcal{H}_1|\mathbf{y}_N) & \textit{if} \quad \mathbf{y}_N \in \mathcal{X}_0 = \mathcal{X} \setminus \mathcal{X}_1 \end{cases}$$

where the data-set for the alternative hypothesis is

$$\mathcal{X}_1 = \left\{ \mathbf{y}_N : \underbrace{\Lambda(\mathbf{y}_N)}_{\log\text{-LR test}} = \log \frac{p(\mathbf{y}_N|\mathcal{H}_1)}{p(\mathbf{y}_N|\mathcal{H}_0)} > 0 \right\}.$$

Tightest upper bound on the minimal error probability

$$P_e^{(N)} = \frac{1}{2}\mathbb{E}\min\left\{p(\mathbf{y}_N|\mathcal{H}_0), p(\mathbf{y}_N|\mathcal{H}_1)\right\} \leq \frac{\exp[-\mu_N(s^\star)]}{2} \leq \frac{\exp[-\mu_N(s)]}{2}$$

where the best exponentially decay rate is obtained for $s^\star \in (0,1)$.

Error Exponent (EE) for Gaussian hypothesis

Thanks to the Stein's lemma :

$$-\lim_{N\to\infty}\frac{\log P_e^{(N)}}{N} = \lim_{N\to\infty, R/N\to c}\frac{\mu_N(s^\star)}{N} \stackrel{\text{def.}}{=} \mu(s^\star)$$

with

$$\mu_N(s) = \frac{1-s}{2}\log\det\left(\mathrm{SNR}\cdot\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \quad (6)$$

$$-\frac{1}{2}\log\det\left(\mathrm{SNR}\cdot(1-s)\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right). \quad (7)$$

The Marcenko-Pastur framework

The Empirical Spectral Measure is $q_N(\lambda) = \frac{1}{N}\sum_{n=1}^{N} \delta(\lambda - \lambda_n(\mathbf{AA}^T))$.

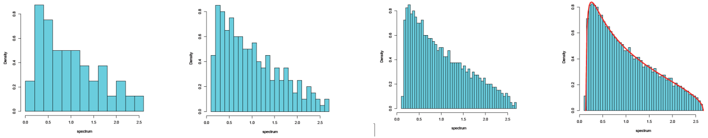
- if R is fixed and $N \to \infty$, then $q_N(\lambda)$ converges toward a non-random measure associated to the Dirac density.
- if $R, N \to \infty$, with $R/N \to c \in (0,1)$ then $q_N(\lambda)$ converges (almost surely) toward a deterministic measure $q_c(\lambda)$ whose density is the well-known Marcenko-Pastur law, *i.e.*,

$$f(\lambda; c) = \frac{\partial q_c(\lambda)}{\partial \lambda} = \frac{\sqrt{(\lambda - \lambda_-)(\lambda_+ - \lambda)}}{2\pi\lambda c} 1_{[\lambda_-, \lambda_+]}(\lambda)$$

where $\lambda_+ = (1+\sqrt{c})^2$ and $\lambda_- = (1-\sqrt{c})^2$.

With $c = 0.7$,

$(40, 100)$ $(200, 500)$ $(800, 2000)$ $(1600, 4000)$



Is it realistic ? YES

- In massive MIMO systems, the large number of sensor (N) allows to improve the resolvability performance, *i.e.*, to resolve more sources (R). Thus, the increase in N implies a larger R in a control way, *i.e.*, with an asymptotic finite ratio.
- In compressed sensing, an universal strategy to verify a concentration inequality is based on random projections, *i.e.*, the dictionary is composed by $\mathbf{\Phi} = \mathbf{S} \cdot \mathbf{B}$ with $\mathbf{B}$ an orthonormal basis and a measurement matrix $\mathbf{S}$ drawn from *i.i.d.* Gaussian entries of zero mean and variance $1/N$. For $0 < \epsilon < 1$, the inequality concentration for dictionary $\mathbf{\Phi}$ is

$$\Pr\left(\left| ||\mathbf{\Phi s}||^2 - ||\mathbf{s}||^2 \right| \geq \epsilon ||\mathbf{s}||^2 \right) \leq \exp[-cN\epsilon^2] \quad (8)$$

where c is a given positive constant.

The state of art and our contribution

Knowing $f(\lambda; c)$, it is possible to derive a closed-form expression for

$$\frac{1}{N} \log\det\left(x \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \rightarrow \int \log(x \cdot \lambda + 1) f(\lambda; c) d\lambda$$

We extend the standard result to our case of interest.

Extension to structured linear matrix

Let $\mathbf{\Phi} = \mathbf{\Phi}_Q \odot \ldots \odot \mathbf{\Phi}_1$

$$\frac{1}{N} \log\det\left(x \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \rightarrow \Psi_c(x)$$

where

$$\Psi_c(x) = \log\left(1 + \frac{2c}{u(x) + (1-c)}\right) + c \cdot \log\left(1 + \frac{2}{u(x) - (1-c)}\right) \quad (9)$$

$$- \frac{4c}{x(u(x)^2 - (1-c)^2)} \text{ with } u(x) = \frac{1}{x} + \sqrt{(\tfrac{1}{x} + \lambda_+)(\tfrac{1}{x} + \lambda_-)}. \quad (10)$$

Application to the EE

Analytical formula

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

$$\frac{\mu_N(s)}{N} \xrightarrow{\text{a.s}} \mu(s) = \frac{(1-s)}{2c}\Psi_c\left(\mathrm{SNR}\right) - \frac{1}{2c}\Psi_c\Big(\mathrm{SNR}\cdot(1-s)\Big). \quad (11)$$

Approximated analytical expressions for $c \ll 1$

The EE can be approximated according to

$$\mu(s) \overset{c\ll 1}{\approx} \frac{c}{2}\Big((1-s)\log(1+\mathrm{SNR}) - \log(1+(1-s)\mathrm{SNR})\Big). \quad (12)$$

The stationary point $s^\star$ is a global minimizer of $\mu(s)$ and is given by

$$s^\star \overset{c\ll 1}{\approx} 1 + \frac{1}{\mathrm{SNR}} - \frac{1}{\log(1+\mathrm{SNR})}. \quad (13)$$

Analysis of $s^\star$ *wrt.* the SNR

We can identify the two following limit scenarios :

- At low SNR, $\mu(s^\star)$, coincides with the EE associated with the BUB based on

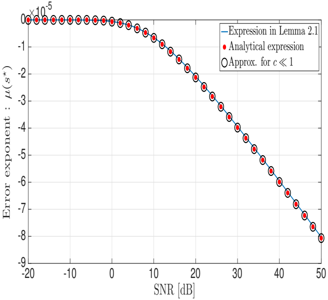
$$\mu\left(\frac{1}{2}\right) \overset{\mathrm{SNR}\ll 1}{\approx} = \frac{c}{2}\log\frac{\sqrt{1+\mathrm{SNR}}}{1+\frac{\mathrm{SNR}}{2}}. \qquad (14)$$

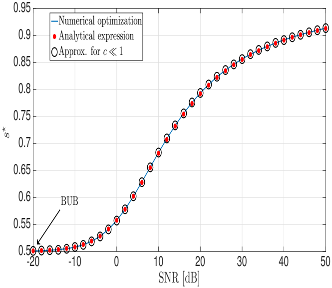
- At contrary for $\mathrm{SNR}\to\infty$, we have $s^\star \to 1$. So, the error exponent associated to BUB cannot be considered as optimal in this regime. Assuming that $\frac{\log \mathrm{SNR}}{\mathrm{SNR}} \to 0$, the EE is given by

$$\mu\left(s^\star\right) \overset{\mathrm{SNR}\ll 1}{\approx} \frac{c}{2}\left(1 - \log\mathrm{SNR} + \log\log(1+\mathrm{SNR})\right). \qquad (15)$$

Numerical illustrations

In this simulation part, $Q = 3$ with $N_1 = N_2 = N_3 = 100$. The factors $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2$ and $\boldsymbol{\Phi}_3$ are generated as a single *i.i.d.* Gaussian realization of rank $R = 20$. We can check $c = 2e - 5 \ll 1$.





Conclusion

1. The derivation and the analyze of the asymptotic performance in terms of minimal Bayes' error probability for the detection of a random parameters is addressed in this work.
2. The term "large" means that the number of available measurements, N, and the number of desired random parameters, R, grow jointly to infinity with an asymptotically constant ratio.
3. The CUB and the error exponent are proposed in closed-form.
4. In addition, it is provided analytical expressions of the optimal parameter s for which the CUB is a tight upper bound on the Bayes' error probability.
5. A main conclusion of this work is that the BUB is the tightest bound at low SNRs. At contrary, this property is no longer true for higher SNRs.