INFORMATION GEOMETRY METRIC FOR RANDOM SIGNAL DETECTION IN LARGE RANDOM SENSING SYSTEMS

Rémy Boyer, Frank Nielsen

HAL Id: hal-01416198

https://hal-centralesupelec.archives-ouvertes.fr/hal-01416198

Submitted on 13 Jul 2017

INFORMATION GEOMETRY METRIC FOR RANDOM SIGNAL DETECTION IN LARGE RANDOM SENSING SYSTEMS

Rémy Boyer

L2S - Department of Signals and Statistics
Université Paris-Sud, CNRS, CentraleSupelec
France
remy.boyer@l2s.centralesupelec.fr

Frank Nielsen

LIX
Ecole Polytechnique
France
nielsen@lix.polytechnique.fr

ABSTRACT

Assume that a N-dimensional noisy measurement vector is available *via* a $N \times R$ linear random sensing operation of a R-dimensional Gaussian signal of interest, denoted by **s**. The problem statement being addressed here is the study of the minimal Bayes' error probability for the detection of **s** where $N \to \infty$ with $N/R \to \beta \in (1, \infty)$. When the exact derivation of this probability is intractable, statistical similarity metrics, nourishing their roots in the information geometry theory, are useful to characterize the exponential rate of the error probability. More precisely, the Chernoff information is asymptotically given by the minimum over $s \in (0, 1)$ of the s-divergence. In many applications, it is hard to evaluate the s-divergence. Worse, due to the asymmetry of the s-divergence for the considered detection problem, the Bhattacharyya divergence ($s = 1/2$), cannot circumvent this problem. As a consequence, the derivation of the optimal value of s requires a costly numerical optimization strategy. In this work, we propose two contributions. The first one is to provide a closed-form expression of the asymptotic normalized s-divergence. The second contribution is to provide an analytic expression for the optimal value of s.

Index Terms— Optimal Bayesian detection, information geometry, minimal error probability, Chernoff/Bhattacharyya upper bound, large random sensing matrix.

1. INTRODUCTION

Evaluate the performance limit for the "Gaussian information plus noise" detection problem is a challenging research topic, see for instance [1–6]. 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 $\Pr(\mathcal{H}_1|\mathbf{y}_N) > \Pr(\mathcal{H}_0|\mathbf{y}_N)$ for a given N-dimensional measurement vector $\mathbf{y}_N$ and 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 [6] but an exact calculation of the minimal Bayes' error probability $P_e^{(N)}$ is often intractable [6,7]. To circumvent this problem, it is standard to exploit well-known information theory based bounds on $P_e^{(N)}$ [8–12]. In particular, the Chernoff information [13,14] is asymptotically (in N) relied to the exponential rate of $P_e^{(N)}$. Due to its strong link to the powerful Amari's divergence in information geometry [14–18], the Chernoff information turns out to be useful in many problems of practical importance as for instance, distributed

sparse detection [19], sparse support recovery [20], energy detection [21], MIMO radar processing [22,23], network secrecy [24], Angular Resolution Limit in array processing [25], detection performance for informed communication systems [26], *etc.* In addition, the Chernoff information can be tight for a maximal s-divergence over parameter $s \in (0, 1)$. Generally, this step requires to solve numerically an optimization problem [27] 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 [8] at the price of a less accurate estimation of $P_e^{(N)}$.

In this work, our primary goal is to derive a closed-form expression of the Chernoff information for any $s \in (0, 1)$ for the detection of a R-dimensional Gaussian signal acquired *via* a $N \times R$ linear random sensing operation corrupted by an additional N-dimensional Gaussian noise for $N \to \infty$ with $N/R \to \beta \in (1, \infty)$. A secondary contribution is to derive a simple approximated analytical expression for the optimal value of s. To reach this goal, it turns out that the large Random Matrix Theory (RMT) framework [28–31] will be relevant.

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 $\Pr(\mathbf{y}_N|\mathcal{H}_i)$ and $\Pr(\mathcal{H}_i|\mathbf{y}_N)$ be the i-th conditional hypothesis and the posterior probabilities, respectively. The Bayes' detection rule chooses the hypothesis $\mathcal{H}_i$ associated with the largest posterior probability $\Pr(\mathcal{H}_i|\mathbf{y}_N)$. Introduce the indicator hypothesis function according to $\phi(\mathbf{y}_N) \sim \text{Bernou}(\alpha)$ where $\text{Bernou}(\alpha)$ stands for the Bernoulli distribution of success probability $\alpha = \Pr(\phi(\mathbf{y}_N) = 1) = \Pr(\mathcal{H}_1)$. Function $\phi(\mathbf{y}_N)$ 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 $\mathcal{X}_0 = \{\mathbf{y}_N : \phi(\mathbf{y}_N) = 0\} = \mathcal{X} \setminus \mathcal{X}_1$ and

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

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

defined as

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

with

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

The standard strategy to minimize $\Pr(\text{Error}|\mathbf{y}_N)$ for a given $\mathbf{y}_N$ is $\min\{\Pr(\mathcal{H}_0|\mathbf{y}_N), \Pr(\mathcal{H}_1|\mathbf{y}_N)\}$ [6]. So using eq. (1), the minimal average error probability can be expressed according to

$$P_e^{(N)} = \mathbb{E}\Big\{ \min\{\Pr(\mathcal{H}_0|\mathbf{y}_N), \Pr(\mathcal{H}_1|\mathbf{y}_N)\} \Big\}$$
$$= \int_{\mathcal{X}} \min\Big\{ (1-\alpha)\Pr(\mathbf{y}_N|\mathcal{H}_0), \alpha\Pr(\mathbf{y}_N|\mathcal{H}_1) \Big\} d\mathbf{y}_N \quad (2)$$

using the Bayes' relation.

2.2. Chernoff Upper Bound (CUB)

Using $\min\{a, b\} \le a^s b^{1-s}$ with $a, b > 0$ and $s \in (0, 1)$ in eq. (2), the minimal error probability is upper bounded according to

$$P_e^{(N)} \le \frac{\alpha}{\tau^s} \int_{\mathcal{X}} \Pr(\mathbf{y}_N|\mathcal{H}_0)^s \Pr(\mathbf{y}_N|\mathcal{H}_1)^{1-s} d\mathbf{y}_N$$
$$\stackrel{\text{def.}}{=} \frac{\alpha}{\tau^s} \cdot \exp[-\mu_N(s)] \qquad (3)$$

where

$$\mu_N(s) = -\log M_{\Lambda(\mathbf{y}_N|\mathcal{H}_1)}(-s) \qquad (4)$$

is the (Chernoff) s-divergence and $M_X(t) = \mathbb{E}\exp[t \cdot X]$ is the moment generating function (mgf) of variable X. Term $\mu_N(s)$ characterizes the exponential rate of the error exponent of $P_e^{(N)}$. The Chernoff information, denoted by $\mu(s)$, is an asymptotic characterization on the best achievable Bayes' error probability and is derived according to [8,10,11,32] :

$$-\lim_{N \to \infty} \frac{\log P_e^{(N)}}{N} = \lim_{N \to \infty} \frac{\mu_N(s)}{N} \stackrel{\text{def.}}{=} \mu(s). \qquad (5)$$

As parameter s is free, the Chernoff information can be retrieved as the unique minimizer:

$$s^\star = \arg\min_{s \in (0,1)} \mu(s). \qquad (6)$$

Finally using eq. (3), eq. (5) and eq. (6), we obtain the Chernoff Upper Bound (CUB). The Bhattacharyya Upper Bound (BUB) is obtained by eq. (3) and eq. (5) and by fixing $s = 1/2$ instead of solving eq. (6).

3. INFORMATION GEOMETRY FOR LARGE RANDOM SENSING MATRICES

3.1. Model definition and associated binary hypothesis test

Assume that we dispose of N noisy measurements collected in vector $\mathbf{y}_N$ *via* a linear random sensing operation where the i.i.d. noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ is assumed to be statistically independent of the i.i.d. signal of interest $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma_\mathbf{s}^2 \mathbf{I}_R)$ where $N > R$. The random sensing system is modeled thanks to the $N \times R$ matrix $\boldsymbol{\Phi}$

whose each entry is an i.i.d. random variable generated as an observable/known single realization of an unspecified probability distribution function parametrized by a zero-mean and a variance of $1/R$. Let $\text{SNR} = \sigma_\mathbf{s}^2/\sigma^2$ be the signal to noise ratio. The detector chooses the null hypothesis $\mathcal{H}_0$ for a null SNR and chooses $\mathcal{H}_1$, otherwise. More formally, the considered binary hypothesis test is

$$\begin{cases} \mathcal{H}_0 & : \quad \mathbf{y}_N \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_0 = \sigma^2 \mathbf{I}_N\right), \\ \mathcal{H}_1 & : \quad \mathbf{y}_N \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_1 = \sigma^2 \left(\text{SNR} \cdot \boldsymbol{\Phi}\boldsymbol{\Phi}^T + \mathbf{I}_N\right)\right). \end{cases} \quad (7)$$

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

$$\mathcal{X}_1 = \left\{\mathbf{y}_N : \mathbf{y}_N^T(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1})\mathbf{y}_N > \tau'\right\} \qquad (8)$$

with $\tau' = \log\frac{\det\boldsymbol{\Sigma}_0}{\tau^2 \det\boldsymbol{\Sigma}_1}$ where $\det(\cdot)$ stands for the determinant. The performance of the above detection problem is clearly related to the question: "How can the covariance matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ be discriminated ? ". This is a canonical problem in the context of the theory of information geometry for covariance matrices.

3.2. Chernoff information for a large sensing matrix

In this section, we derive a closed-form expression the CUB for the test of eq. (7) in the regime $N \to \infty$ with $N/R \to \beta \in (1, \infty)$.

Lemma 3.1 *The log-mgf given by eq.* (4) *for test of eq.* (7) *is given by*

$$\mu_N(s) = \frac{1-s}{2} \log\det\left(\text{SNR} \cdot \boldsymbol{\Phi}\boldsymbol{\Phi}^T + \mathbf{I}\right)$$
$$- \frac{1}{2} \log\det\left(\text{SNR} \cdot (1-s)\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \mathbf{I}\right). \qquad (9)$$

Proof See Appendix 6.1

Result 3.2 *Using Lemma 3.1, the Chernoff information defined in eq.* (5) *takes the following simple expression:*

$$\frac{\mu_N(s)}{N} \xrightarrow{\text{a.s}} \mu(s)$$
$$= \frac{(1-s)}{2\beta} \Psi_\beta(\text{SNR}) - \frac{1}{2\beta} \Psi_\beta\left(\text{SNR} \cdot (1-s)\right) \quad (10)$$

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

$$\Psi_\beta(x) = \beta \log\left(1 + x - \frac{\Phi_\beta^2(x)}{4}\right) + \log\left(1 + \beta x - \frac{\Phi_\beta^2(x)}{4}\right)$$
$$- \frac{\Phi_\beta^2(x)}{4x}, \quad \Phi_\beta(x) = \sqrt{x(1 + \sqrt{\beta})^2 + 1} - \sqrt{x(1 - \sqrt{\beta})^2 + 1}.$$

Proof See Appendix 6.2.

Using Result 3.2, the Bhattacharyya information is given by

$$\mu\left(\frac{1}{2}\right) = \frac{\Psi_\beta(\text{SNR})}{4\beta} - \frac{1}{2\beta} \Psi_\beta\left(\frac{\text{SNR}}{2}\right). \qquad (11)$$

3.3. Chernoff information in the high SNR regime

The derivation of an analytic expression for $\mu(s)$ allows to obtain for an insignificant computational cost the CUB. This is crucial for large sensing systems. Another advantage is that given an analytic expression for $\mu(s)$, it is possible to derive in closed-form $s^\star$ and thus to avoid the costly numerical optimization step of eq. (6). The following Results provide analytic expressions for $s^\star$ and for the Chernoff information in the high SNR regime.

Result 3.3 *In the high SNR regime, the optimal parameter s is obtained analytically according to*

$$s^\star \overset{\mathrm{SNR}\gg 1}{\approx} 1 - \frac{1}{\Psi_\beta(\mathrm{SNR})}. \tag{12}$$

Proof See Appendix 6.3.

Result 3.4 *The Chernoff information takes the following simple expression:*

$$\mu(s^\star) \overset{\mathrm{SNR}\gg 1}{\approx} \frac{1}{2\beta}\left(1 - \Psi_\beta\left(\frac{\mathrm{SNR}}{\Psi_\beta(\mathrm{SNR})}\right)\right). \tag{13}$$

Proof The proof is straightforward using eq. (10) and eq. (12). □

4. NUMERICAL ILLUSTRATIONS

In this simulation part, the sensing matrix $\mathbf{\Phi}$ is generated as a prescribed single 60×10 Gaussian realization with $\mathbb{E}[\mathbf{\Phi}]_{n,r} = 0$ and $\mathbb{E}[\mathbf{\Phi}]^2_{n,r} = \frac{1}{R}\delta_{n-r}$. We fix $\alpha = 1/2$. On Fig. 1, the Chernoff information (using Result 3.2) and the normalized s-divergence with respect to parameter s are drawn. First, we can see that the numerical and closed-form expressions are almost merged. In addition, the Chernoff information is asymmetric. This means that $s^\star \neq 1/2$ and thus the Bhattacharyya information cannot be considered as the best strategy. On Fig. 2, the closed-form expression of eq. (12) obtained in the high SNR regime is illustrated by comparison with the numerical optimization of eq. (6) involving eq. (9). We can note that the proposed analytical expression of eq. (12) accurately predicts the coefficient $s^\star$ for a wide range of SNR greater than approximatively 10 dB. We can also note that the BUB is interesting only in the very low SNR regime. To study the detection performance, the upper bound on $P_e^{(N)}$ has been drawn in linear (*resp.* log) scale on Fig. 3 (*resp.* Fig. 4) for a wide range of SNRs. On these figures, several configurations of CUB and BUB are compared and detailed in Table 1. We can check that the BUB_2 given by eq. (11) as a closed-form expression is an efficient and low cost upper bound for negative SNRs. Fig. 3 and in particular Fig. 4 illustrate that the BUB is a relaxed upper bound in the high SNR regime. As expected, the CUB_3 which is derived in the high SNR regime is not informative for negative SNRs. But, we can observe that the CUB_3 given by eq. (13) provides a very accurate and low cost solution to upper bound the error probability $P_e^{(N)}$ for positive SNR.

5. CONCLUSION

In the Bayes' detection framework, the performance in term of the minimal error probability, $P_e^{(N)}$, for the detection of a randomly acquired R-dimensional Gaussian signal corrupted by a N-dimensional Gaussian noise is a canonical problem in the framework of the geometry information of covariance matrices. Indeed, $P_e^{(N)}$
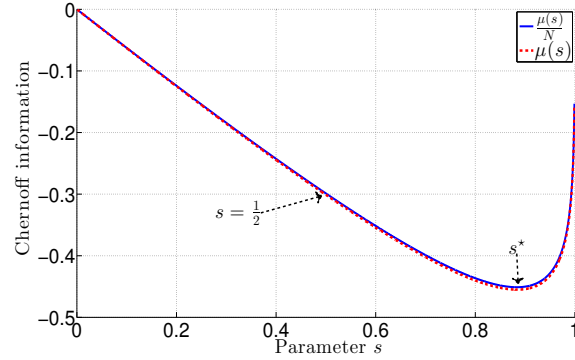

Fig. 1. Chernoff information and normalized s-divergence *vs* s with $\mathrm{SNR} = 20$ dB
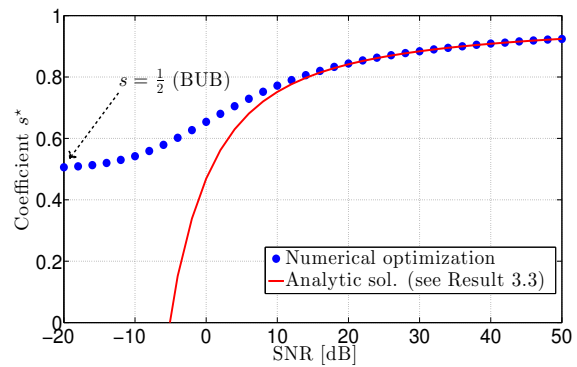

Fig. 2. Optimal value of parameter s *vs* SNR in dB

admits asymptotically an upper bound governed by the Chernoff information. In many applications, it is hard to obtain closed-form expressions of the s-divergence for any $s \in (0, 1)$ and of the optimal parameter $s^\star$ in a realistic computational time. The default choice $s = 1/2$ corresponding to the Bhattacharyya information is shown to be not the optimal strategy since the s-divergence is asymmetric for the detection problem of interest. In this work, we propose two contributions. The first one is to provide a closed-form expression of the Chernoff information for linear random sensing systems. The derivation of such analytic expressions has two mains advantages. Firstly, the Chernoff information for any $s \in (0, 1)$ and Bhattacharyya information can be evaluated thanks to an analytic expression and thus for a very low computational cost. Secondly, the

	Normalized log-mgf	$s^\star$	Comput. cost
CUB_1	Numeric with eq. (9)	Numeric	Highest
CUB_2	Analytic with eq. (10)	Numeric	Medium
CUB_3	Analytic with eq. (13)	Analytic	Lowest
BUB_1	Numeric with eq. (9)	$1/2$	High
BUB_2	Analytic with eq. (11)	$1/2$	Lowest

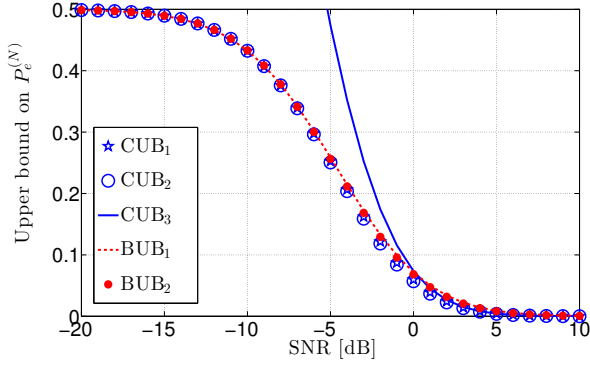**Table 1**. Description of the tested configurations in Fig. 3 and 4

Fig. 3. Upper bound on $P_e^{(N)}$ in linear scale *Vs.* SNR in dB. The acronyms are detailed in Table 1
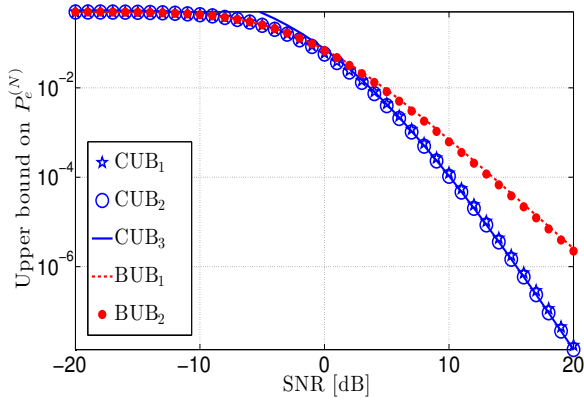

Fig. 4. Upper bound on $P_e^{(N)}$ in log-scale *Vs.* SNR in dB. The acronyms are detailed in Table 1

optimal parameter s is derived analytically in the high SNR regime.

6. APPENDIX

6.1. Proof of Lemma 3.1

The log-mgf in eq. (4) for the following binary hypothesis test

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

is given by [14]:

$$\mu_N(s) = \frac{1}{2} \log \frac{\det(s\mathbf{\Sigma}_0 + (1-s)\mathbf{\Sigma}_1)}{[\det\mathbf{\Sigma}_0]^s[\det\mathbf{\Sigma}_1]^{1-s}}. \tag{14}$$

Using the expressions of the covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, the numerator in eq. (14) is given by

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

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

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

Using the above expressions, $\mu_N(s)$ is given by eq. (9).

6.2. Proof of Result 3.2

For $N, R \to \infty$ with $N/R \to \beta \in (1, \infty)$, eq. (10) is derived using Lemma 3.1 and the following property (see [28–31] for instance): $\frac{1}{N} \log \det \left(x \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right) \xrightarrow{\text{a.s}} \frac{1}{\beta} \Psi_\beta(x)$ where function $\Psi_\beta(\cdot)$ is defined in Result 3.2.

6.3. Proof of Result 3.3

The first step of the proof is based on the derivation of an alternative expression of $\mu_N(s)$ given by eq. (14) involving the inverse of the covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$. Specifically, we have

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

$$= -\frac{1}{2} \log \frac{\det\left([(1-s)\mathbf{\Sigma}_0^{-1} + s\mathbf{\Sigma}_1^{-1}]^{-1}\right)}{[\det\mathbf{\Sigma}_0]^{1-s}[\det\mathbf{\Sigma}_1]^s}. \tag{15}$$

The second step is to derive a closed-form expression in the high SNR regime using the following the approximation (see [33] for instance): $\left(x \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right)^{-1} \overset{x \gg 1}{\approx} \mathbf{\Pi}_\mathbf{\Phi}^\perp = \mathbf{I}_N - \mathbf{\Phi}\mathbf{\Phi}^\dagger$ where $\mathbf{\Pi}_\mathbf{\Phi}^\perp$ is an orthogonal projector such as $\mathbf{\Pi}_\mathbf{\Phi}^\perp \mathbf{\Phi} = \mathbf{0}$ and $\mathbf{\Phi}^\dagger = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T$. The numerator in eq. (15) is given by

$$\left[(1-s)\mathbf{\Sigma}_0^{-1} + s\mathbf{\Sigma}_1^{-1}\right]^{-1} \overset{\text{SNR} \gg 1}{\approx} \sigma^2 \left(\mathbf{I}_N - s\mathbf{I}_N + s\mathbf{\Pi}_\mathbf{\Phi}^\perp\right)^{-1}$$

$$= \sigma^2 \left(\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^\dagger\right)^{-1}.$$

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

$$\lambda_n \left\{\left(\mathbf{I}_N - s\mathbf{\Phi}\mathbf{\Phi}^\dagger\right)^{-1}\right\} = \frac{1}{\lambda_n\{\mathbf{I}_N\} - s\lambda_n\{\mathbf{\Phi}\mathbf{\Phi}^\dagger\}}$$

$$= \begin{cases} \frac{1}{1-s}, & 1 \le n \le R, \\ 1, & R+1 \le n \le N \end{cases}$$

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

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

$$= -R \log(1-s).$$

Finally, thanks to eq. (15), we have

$$\frac{\mu_N(s)}{N} \overset{\text{SNR} \gg 1}{\approx} \frac{R}{2N} \log(1-s) + \frac{s}{2N} \log \det \left(\text{SNR} \cdot \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{I}\right)$$

$$\xrightarrow{\text{a.s}} \mu(s) \overset{\text{SNR} \gg 1}{\approx} \frac{1}{2\beta}\left(\log(1-s) + s \cdot \Psi_\beta(\text{SNR})\right).$$

Finally, to obtain $s^\star$ in eq. (12), we solve $\frac{\partial \mu(s)}{\partial s} = 0$.

7. REFERENCES

[1] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *Journal of Multivariate Analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.

[2] P. Loubaton and P. Vallet, "Almost sure localization of the eigenvalues in a gaussian information plus noise model–application to the spiked models," *Electron. J. Probab*, vol. 16, pp. 1934–1959, 2011.

[3] P. Bianchi, M. Debbah, M. Maïda, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Transactions on Information theory*, vol. 57, no. 4, pp. 2400–2419, 2011.

[4] J. W. Silverstein and P. L. Combettes, "Signal detection via spectral theory of large dimensional random matrices," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 2100–2105, 1992.

[5] O. Besson, L. L. Scharf, and F. Vincent, "Matched direction detectors and estimators for array processing with subspace steering vector uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4453–4463, 2005.

[6] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory, vol. 1*. Prentice Hall Upper Saddle River, NJ, USA:, 1998.

[7] H. Avi-Itzhak and T. Diep, "Arbitrarily tight upper and lower bounds on the Bayesian probability of error," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 89–91, Jan 1996.

[8] S. Sinanović and D. H. Johnson, "Toward a theory of information processing," *signal processing*, vol. 87, no. 6, pp. 1326–1344, 2007.

[9] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE transactions on communication technology*, vol. 15, no. 1, pp. 52–60, 1967.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[11] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 131–142, 1966.

[12] M. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 368–372, Jul 1970.

[13] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, pp. 493–507, 1952.

[14] F. Nielsen, "Chernoff information of exponential families," *CoRR*, vol. abs/1102.2684, 2011. [Online]. Available: http://arxiv.org/abs/1102.2684

[15] S.-I. Amari and H. Nagaoka, *Methods of information geometry*. American Mathematical Soc., 2007, vol. 191.

[16] F. Nielsen and R. Bhatia, *Matrix information geometry*. Springer, 2013.

[17] L. Yang, M. Arnaudon, and F. Barbaresco, "Riemannian median, geometry of covariance matrices and radar target detection," in *European Radar Conference (EuRAD)*. IEEE, 2010, pp. 415–418.

[18] P. Formont, J.-P. Ovarlez, and F. Pascal, "On the use of matrix information geometry for polarimetric sar image classification," in *Matrix Information Geometry*. Springer, 2013, pp. 257–276.

[19] S. P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1446–1460, 2015.

[20] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE transactions on information theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[21] Y. Lee and Y. Sung, "Generalized Chernoff information for mismatched Bayesian detection and its application to energy detection," *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 753–756, Nov 2012.

[22] S. Sen and A. Nehorai, "OFDM MIMO radar with mutual-information waveform design for low-grazing angle tracking," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3152–3162, 2010.

[23] E. Grossi and M. Lops, "Space-time code design for MIMO detection based on Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3989–4004, 2012.

[24] R. Boyer and C. Delpha, "Relative-entropy based beamforming for secret key transmission," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2012, pp. 265–268.

[25] N. D. Tran, R. Boyer, S. Marcos, and P. Larzabal, "Angular Resolution Limit for array processing: Estimation and information theory approaches," in *20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 101–105.

[26] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "Joint estimation and detection against independence," in *Annual Conference on Communication, Control, and Computing (Allerton)*, Sept 2014, pp. 1220–1227.

[27] F. Nielsen, "An information-geometric characterization of Chernoff information," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 269–272, 2013.

[28] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA, USA: Now Publishers Inc., Jun. 2004, vol. 1, no. 1.

[29] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.

[30] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Sbornik: Mathematics*, vol. 1, no. 4, pp. 457–483, 1967.

[31] J. W. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[32] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.

[33] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1413–1424, 1994.