# On the second-order statistics of the EVD of sample covariance matrices : application to the detection of noncircular or/and nonGaussian components

Jean-Pierre Delmas, Yann Meurisse

# On the second-order statistics of the EVD of sample covariance matrices - Application to the detection of noncircular or/and nonGaussian components

Jean Pierre Delmas and Yann Meurisse

**Abstract**

This paper presents an asymptotic analysis of the eigen value decomposition (EVD) of the sample covariance matrix associated with independent identically distributed (IID) non necessarily circular and Gaussian data that extends the well known analysis presented in the literature for circular and Gaussian data. Closed-form expressions of the asymptotic bias and variance of the sample eigenvalues and eigenvectors are given. As an application of these extended expressions, the statistical performance analysis of the widely-used minimum description length (MDL) criterion applied to the detection of the number of noncircular or/and nonGaussian sources impinging on an array of sensors is considered with a particular attention paid to uncorrelated rectilinear sources.

**Index Terms**

Eigen value decomposition (EVD), sample covariance matrix, eigenvalue, eigenvector, minimum description length (MDL), source detection, asymptotic performance analysis, noncircular, nonGaussian, rectilinear sources.

Revised correspondance accepted to IEEE Transactions on Signal Processing

## I. Introduction

Eigenvalues and eigenvectors of sample covariance matrices are used in the solution of a wide range of statistical signal processing problems, in particular in spectral analysis and array processing among many others. The first and second-order statistics of this EVD are needed to assess the performance in terms of bias/variance

Jean Pierre Delmas and Yann Meurisse are with TELECOM SudParis, Département CITI, CNRS UMR 5157, 91011 Evry Cedex, France, e-mail:jean-pierre.delmas@it-sudparis.eu, yann.meurisse@wanadoo.fr, phone: +(33).1.60.76.46.32, fax: +(33).1.60.76.44.33.

of estimators or probability of events derived from these eigenvalues and eigenvectors. The statistics of the sample eigenvalues and eigenvectors have been widely studied in the statistical literature (e.g., [1],[2]) and are quoted in standard texts such as [3] and [4]. The tricky issue of the uniqueness of the eigenvectors for complex-valued data has been considered in [5] and then in [6]. But to the best of our knowledge, all the published results has been derived only under the assumption of real-valued or circular complex-valued Gaussian distributions of the data.

The main aim of this paper is to extend these results to arbitrary real or complex fourth-order distributions of the data, where closed-form expressions of asymptotic bias and variance of the sample eigenvalues and eigenvectors are derived. As an application of these extended expressions, the statistical performance analysis of the widely-used MDL criterion introduced by Rissanen [7] and popularized by Wax and Kailath [8], is considered to the detection of the number of noncircular or/and nonGaussian sources impinging on an array of sensors. Note that optimal detection (in the sense of maximum likelihood under the Gaussian assumption) of the number of circular and noncircular sources has been recently studied in [9]. But naturally, these extended expressions may find other applications in multivariate analysis. We focus here on the probability of underestimating the number of sources in the case of a single or two sources under asymptotic conditions (with respect to the number of snapshots) and around the threshold regions, following the approach proposed in [10],[11], and then recently improved in [12]. We show in particular that the numerical values of this probability of underestimating the number of sources given by the MDL detector derived from the standard sample covariance matrix are not robust to the noncircularity and/or the nonGaussianity of the data. We prove that this probability decreases for uncorrelated[1] rectilinear[2] sources when the MDL detector is derived from the augmented sample covariance matrix of the data with respect to the standard one. Furthermore, when the sources of fixed DOAs are equipowered, this probability is minimized for in quadrature complex envelopes of these sources. This extends to the detection, the well known performance in terms of variance [14] and resolving power [15] of the estimated directions of arrival (DOA) of uncorrelated rectilinear sources.

The paper is organized as follows. In Section II, the asymptotic statistics of the EVD of sample covariance matrices is addressed with a particular attention paid to the determination of the selected eigenvectors for complex-valued data. Section III applies the closed-form expressions of the asymptotic bias and variance of the sample eigenvalues to the performance analysis of the MDL detection of the number of components of a linear model. Section IV specializes these results to the number of noncircular or/and nonGaussian sources impinging on an array of sensors and gives some numerical illustrations compared to Monte Carlo experiments.

[1]For general complex-valued random variables (RVs), RVs are called uncorrelated if the real and imaginary parts of these RVs are uncorrelated [13]. A necessary and suffisant condition to be uncorrelated is that both the associated covariance and the complementary covariance matrices are diagonal.

[2]A scalar complex-valued RV is called rectilinear or maximally improper (term used by other researchers), if the support of its distribution collapses to a line in the complex plane.

The following notations are used throughout the paper. Matrices and vectors are represented by bold upper case and bold lower case characters, respectively. Vectors are by default in column orientation, while $T$, $H$ and $*$ stand for transpose, conjugate transpose, conjugate respectively. $\iota = \sqrt{-1}$ is the imaginary unit. $\mathrm{vec}(\cdot)$ is the "vectorization" operator that turns a matrix into a vector by stacking the columns of the matrix one below another. The Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the block matrix whose $(i,j)$ block element is $a_{i,j}\mathbf{B}$ and the vec-permutation matrix $\mathbf{K}$ transforms $\mathrm{vec}(\mathbf{C})$ to $\mathrm{vec}(\mathbf{C}^T)$ for any matrix $\mathbf{C}$.

## II. Asymptotic statistics of EVD of sample covariance matrices

Consider a sequence $\mathbf{x}_{t=1,..,T} \in \mathbb{C}^n$ of IID zero-mean complex multidimensional random variables (RV) with finite fourth-order moments. The standard covariance matrix, the complementary covariance matrix [16] and the quadrivariance matrix of $\mathbf{x}_t$ are respectively given by $\mathbf{R} \overset{\text{def}}{=} \mathrm{E}(\mathbf{x}_t\mathbf{x}_t^H)$, $\mathbf{R}' \overset{\text{def}}{=} \mathrm{E}(\mathbf{x}_t\mathbf{x}_t^T)$ and $(\mathbf{Q})_{i+(j-1)n,l+(k-1)n} \overset{\text{def}}{=} \mathrm{Cum}(x_{t,i}, x_{t,j}^*, x_{t,k}, x_{t,l}^*)$ where $\mathbf{x}_t = (x_{t,1}, x_{t,2}, ..., x_{t,n})^T$. The sample covariance estimate of $\mathbf{R}$ is usually defined as $\widehat{\mathbf{R}} \overset{\text{def}}{=} \frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t^H = \mathbf{R} + \delta\mathbf{R}$ which can be considered as a perturbation of $\mathbf{R}$. To derive the asymptotic distribution of the EVD of $\widehat{\mathbf{R}}$, we need the following extension of an identity derived under the circular Gaussian distribution of the data in [4, p.114].

*Lemma 1:* For arbitrary vectors $(\mathbf{a}_i)_{i=1,...,4} \in \mathbb{C}^n$ and distributions with finite fourth-order moment of $\mathbf{x}_t$ not necessarily circular and Gaussian, the following identity is proved in the Appendix.

$$\mathrm{E}[(\mathbf{a}_1^H\delta\mathbf{R}\mathbf{a}_2)(\mathbf{a}_3^H\delta\mathbf{R}\mathbf{a}_4)] = \frac{1}{T}\left\{(\mathbf{a}_1^H\mathbf{R}\mathbf{a}_4)(\mathbf{a}_3^H\mathbf{R}\mathbf{a}_2) + (\mathbf{a}_1^H\mathbf{R}'\mathbf{a}_3^*)(\mathbf{a}_2^T\mathbf{R}'^*\mathbf{a}_4) + (\mathbf{a}_2^T \otimes \mathbf{a}_1^H)\mathbf{Q}(\mathbf{a}_3^* \otimes \mathbf{a}_4)\right\}. \quad (1)$$

We assume that the eigenvalues $(\lambda_i)_{i=1,..,n}$ of $\mathbf{R}$, ordered in decreasing order satisfy the condition $\lambda_1 > ... > \lambda_r > \lambda_{r+1} = ... = \lambda_n = \sigma^2$. Let $(\mathbf{v}_i)_{i=1,..,n}$ be an arbitrary set of associated orthonormal eigenvectors. We note that $(\mathbf{v}_i)_{i=r+1,..,n}$ are defined up to an arbitrary unitary transformation, in contrast to $(\mathbf{v}_i)_{i=1,..,r}$ that are defined up to multiplicative unit modulus complex number, but which are arbitrarily fixed[3].

Consider now for a "small enough" perturbation term $\delta\mathbf{R}$ of $\mathbf{R}$, the EVD of $\widehat{\mathbf{R}} = \mathbf{R} + \delta\mathbf{R}$

$$\widehat{\mathbf{R}}\bar{\mathbf{v}}_i = \widehat{\lambda}_i\bar{\mathbf{v}}_i, \quad i = 1, ..., n, \quad (2)$$

where $(\widehat{\lambda}_i)_{i=1,...,n}$ denotes the perturbation of $\lambda_i$ that satisfies $\widehat{\lambda}_1 > \widehat{\lambda}_2 > ... > \widehat{\lambda}_n$, and where $(\bar{\mathbf{v}}_i)_{i=1,...,r}$ are the associated eigenvectors uniquely determined from $(\mathbf{v}_i)_{i=1,...,r}$ by

$$\bar{\mathbf{v}}_i^H\mathbf{v}_i = 1, \quad i = 1, ..., r. \quad (3)$$

We consider in the following the eigenvectors $(\widehat{\mathbf{v}}_i)_{i=1,...,r}$ defined by $\widehat{\mathbf{v}}_i = \bar{\mathbf{v}}_i/||\bar{\mathbf{v}}_i||$. We note that in contrast to the determination of $\mathbf{v}_i$ which is arbitrary fixed, the determination of $\widehat{\mathbf{v}}_i$ must be related to the choice of

[3]For example, the MATLAB's svd function produces a set of singular vectors (which can be used in lieu of eigenvectors for the Hermitian covariance matrix $\mathbf{R}$) that are orthonormal and have a real first entry.

$\mathbf{v}_i$ to consider bias, variance and covariance of $(\widehat{\mathbf{v}}_i)_{i=1,...,r}$ in Result 1. Note that an alternative solution would be to choose for both $\mathbf{v}_i$ and $\widehat{\mathbf{v}}_i$, the MATLAB's normalization. The second-order statistics associated with this determination has been partially studied in [6], where very complicated expressions of $\mathrm{Cov}(\widehat{\mathbf{v}}_i, \widehat{\mathbf{v}}_j)$ and $\mathrm{Cov}(\widehat{\mathbf{v}}_i, \widehat{\mathbf{v}}_j^*)$ are given [6, rels (37-38)] for zero-mean circular Gaussian distributions. We have not opted for this determination because of its complexity.

We are now interested by the asymptotic distribution of $(\widehat{\lambda}_1, ..., \widehat{\lambda}_r, \widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_r)$ with respect to the number $T$ of data. We adopt a functional analysis that consists of recognizing that the whole process of constructing the estimate $(\widehat{\lambda}_1, ..., \widehat{\lambda}_r, \widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_r)$ from $\widehat{\mathbf{R}}$ is equivalent to defining an infinitely differentiable mapping [17] on a neighborhood of $(\lambda_1, ..., \lambda_r, \mathbf{v}_1, ..., \mathbf{v}_r)$ linking $(\widehat{\lambda}_1, ..., \widehat{\lambda}_r, \widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_r)$ to the statistics $\widehat{\mathbf{R}}$ from which it is inferred: $\widehat{\mathbf{R}} \longmapsto (\widehat{\lambda}_1, ..., \widehat{\lambda}_r, \widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_r) = g(\widehat{\mathbf{R}})$. Using the central limit theorem applied to the IID RVs $\mathrm{vec}(\mathbf{x}_t\mathbf{x}_t^H) = \mathbf{x}_t^* \otimes \mathbf{x}_t$, we have[4] $\sqrt{T}\ \left(\mathrm{vec}(\widehat{\mathbf{R}}) - \mathrm{vec}(\mathbf{R})\right) \xrightarrow{\mathcal{L}} \mathcal{N}_C\left(\mathbf{0}; \mathbf{C}_R, \mathbf{C}'_R\right)$, where [18]

$$\mathbf{C}_R = (\mathbf{R}^* \otimes \mathbf{R}) + \mathbf{K}(\mathbf{R}' \otimes \mathbf{R}'^*) + \mathbf{Q} \quad \text{and} \quad \mathbf{C}'_R = \mathbf{C}_R\mathbf{K} \tag{4}$$

and the standard theorem of continuity[5] (see e.g., [19, p.122]), allows us to deduce that the estimates $(\widehat{\lambda}_1, ..., \widehat{\lambda}_r, \widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_r)$ are likewise asymptotically Gaussian distributed with the same convergence speed $\sqrt{T}$.

The first and second-order statistics of this asymptotic distribution are deduced from a second-order Taylor expansion of $g(.)$ at point $\mathbf{R}$, where the remainder issue can be dealt with rigourously and conveniently [20]. This approach called delta method is used in practice from a second-order perturbation analysis that is developed in the Appendix. This allows us to prove the following result

*Result 1:* The asymptotic first and second-order statistics of the estimates $(\widehat{\lambda}_1, ..., \widehat{\lambda}_r, \widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_r)$ for arbitrary distributions with finite fourth-order moment of $\mathbf{x}_t$ not necessarily circular and Gaussian are given by

$$\mathrm{E}(\widehat{\lambda}_i) = \lambda_i + \frac{1}{T} \sum_{1 \leq k \neq i \leq n} \frac{\lambda_i\lambda_k + |\lambda_{i,k}|^2 + \lambda_{i,k,i,k}}{\lambda_i - \lambda_k} + o(\frac{1}{T}) \tag{5}$$

$$\mathrm{Cov}(\widehat{\lambda}_i, \widehat{\lambda}_j) = \frac{1}{T}\left(\lambda_i^2\delta_{i,j} + |\lambda_{i,j}|^2 + \lambda_{i,i,j,j}\right) + o(\frac{1}{T}) \tag{6}$$

$$\begin{aligned} \mathrm{E}(\widehat{\mathbf{v}}_i) &= \left(1 - \frac{1}{2T} \sum_{1 \leq k \neq i \leq n} \frac{\lambda_i\lambda_k + |\lambda_{i,k}|^2 + \lambda_{i,k,i,k}}{(\lambda_i - \lambda_k)^2}\right) \mathbf{v}_i \\ &+ \frac{1}{T} \sum_{1 \leq k \neq i \leq n} \left( -\frac{\lambda_{i,k}\lambda_{i,i}^* + \lambda_{i,i,k,i}}{(\lambda_i - \lambda_k)^2} + \sum_{1 \leq l \neq i \leq n} \frac{\lambda_{k,l}\lambda_{i,l}^* + \lambda_{l,k,l,i}}{(\lambda_i - \lambda_k)(\lambda_i - \lambda_l)} \right) \mathbf{v}_k + o(\frac{1}{T}) \end{aligned} \tag{7}$$

[4] $\xrightarrow{\mathcal{L}}$ means the convergence in distribution when $T \to \infty$, while $\mathcal{N}_C(\mathbf{m}, \mathbf{C}, \mathbf{C}')$ denotes the complex Gaussian distribution whose mean, covariance and complementary covariance are $\mathbf{m}$, $\mathbf{C}$ and $\mathbf{C}'$, respectively.

[5] We state this theorem for the convenience of the readers. Suppose that a vector complex-valued sequence $\mathbf{x}_T$ satisfies $\sqrt{T}(\mathbf{x}_T - \mathbf{m}) \xrightarrow{\mathcal{L}} \mathcal{N}_C(\mathbf{0}; \mathbf{C}_x, \mathbf{C}'_x)$. Let $\mathbf{y}_T = g(\mathbf{x}_T)$ be a differentiable function with a nonzero differential $\mathbf{D} = \left[\frac{\partial g_i}{\partial x_j}\right]_{\mathbf{x}=\mathbf{m}}$. Then $\sqrt{T}(g(\mathbf{x}_T) - g(\mathbf{m})) \xrightarrow{\mathcal{L}} \mathcal{N}_C(\mathbf{0}; \mathbf{C}_y, \mathbf{C}'_y)$ with $\mathbf{C}_y = \mathbf{D}\mathbf{C}_x\mathbf{D}^H$ and $\mathbf{C}'_y = \mathbf{D}\mathbf{C}'_x\mathbf{D}^T$.

$$\mathrm{Cov}(\widehat{\mathbf{v}}_i,\widehat{\mathbf{v}}_j) = \frac{\delta_{i,j}}{T}\sum_{1\leq k\neq i\leq n}\frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\mathbf{v}_k\mathbf{v}_k^H + \frac{1}{T}\sum_{1\leq k\neq i\leq n}\sum_{1\leq l\neq j\leq n}\frac{\lambda_{j,k}\lambda_{i,l}^* + \lambda_{i,k,j,kl}}{(\lambda_i-\lambda_k)(\lambda_j-\lambda_l)}\mathbf{v}_k\mathbf{v}_l^H + o(\frac{1}{T}) \quad (8)$$

$$\mathrm{Cov}(\widehat{\mathbf{v}}_i,\widehat{\mathbf{v}}_j^*) = -\frac{(1-\delta_{i,j})}{T}\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\mathbf{v}_j\mathbf{v}_i^T + \frac{1}{T}\sum_{1\leq k\neq i\leq n}\sum_{1\leq l\neq j\leq n}\frac{\lambda_{k,l}\lambda_{i,j}^* + \lambda_{i,k,l,j}}{(\lambda_i-\lambda_k)(\lambda_j-\lambda_l)}\mathbf{v}_k\mathbf{v}_l^T + o(\frac{1}{T}) \quad (9)$$

$$\mathrm{Cov}(\widehat{\lambda}_i,\widehat{\mathbf{v}}_j) = \frac{1}{T}\sum_{1\leq k\neq j\leq n}\left(\frac{\lambda_{i,k}\lambda_{i,j}^* + \lambda_{i,i,k,j}}{\lambda_j-\lambda_k}\right)\mathbf{v}_k + o(\frac{1}{T}), \quad (10)$$

for $i,j=1,...r$ and where $\mathrm{Cov}(\mathbf{z}_1,\mathbf{z}_2) \overset{\mathrm{def}}{=} \mathrm{E}\left((\mathbf{z}_1-\mathrm{E}(\mathbf{z}_1))(\mathbf{z}_2-\mathrm{E}(\mathbf{z}_2))^H\right)$, $\delta_{i,j}$ is the Kronecker delta, $\lambda_{i,j} \overset{\mathrm{def}}{=} \mathbf{v}_i^H\mathbf{R}'\mathbf{v}_j^*$ and $\lambda_{i,j,k,l} \overset{\mathrm{def}}{=} (\mathbf{v}_i^T\otimes\mathbf{v}_j^H)\mathbf{Q}(\mathbf{v}_k^*\otimes\mathbf{v}_l)$.

**Remark 1:** Naturally all the expressions of this result reduce to the first and second-order statistics given in [4, Th.9.2.4] (where the determination of $\widehat{\mathbf{v}}_i$ is not specified) and in [5] (where determination (3) is used) for the circular Gaussian distribution of $\mathbf{x}_t$ for which $\lambda_{i,j}=0$ and $\lambda_{i,j,k,l}=0$.

**Remark 2:** In contrast to the circular Gaussian distribution, the estimated eigenvalues are no longer asymptotically independent between each others (6) and the estimated eigenvalues and eigenvectors are no longer asymptotically independent for arbitrary distributions of $\mathbf{x}_t$ (10).

**Remark 3:** We note that the asymptotic distribution of the estimated eigenvectors is sensitive to the noncircularity and nonGaussiannity of $\mathbf{x}_t$, in contrast to the asymptotic distribution of the estimated principal projector, $\widehat{\mathbf{\Pi}} \overset{\mathrm{def}}{=} \sum_{i=1}^r \widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^H$ that does not depend on the distribution of $\mathbf{x}_t$ [21].

**Remark 4:** $\lambda_{i,j,k,l}$ is generally complex-valued, but it is straightforward to prove from (4) that $\mathbf{Q}$ is Hermitian and that $(\mathbf{a}^T\otimes\mathbf{a}^H)\mathbf{Q}(\mathbf{b}^*\otimes\mathbf{b})$ is real-valued for arbitrary vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{C}^n$, so $\lambda_{i,j,i,j}$ and $\lambda_{i,i,j,j}$ that appear in (5) and (6) are real-valued.

**Remark 5:** Note that Lemma 1 is still valid for real-valued data $\mathbf{x}_t$ and vectors $(\mathbf{a}_i)_{i=1,...,4}$ replacing $\mathbf{R}'$ by $\mathbf{R}$, in contrast to Result 1, for which the different expressions of the first and second-order statistics that can be derived using the same approach, are slightly different.

**Remark 6:** Note that in contrast to theoretical Result 1, which is valid for arbitrary eigenvalues that satisfy $\lambda_1>...\lambda_r>\lambda_{r+1}=...=\sigma^2$, the approximations deduced from Result 1 must take into account the separation between two successive eigenvalues which cannot be "too small", as an example would be given in Section IV.

## III. Application to the detection of noncircular and/or nonGaussian components

The MDL criterion is one of the most successful information theoretic criteria for estimating the number $r$ of components of

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{n}_t \quad (11)$$

where $\mathbf{s}_t=(s_{t,1},...,s_{t,r})$, $\mathrm{E}(\mathbf{s}_t\mathbf{s}_t^H)$ is not singular, $\mathbf{A}$ is an $n\times r$ full column rank matrix with $r<n$ and $\mathbf{s}_t$ and $\mathbf{n}_t$ are uncorrelated with $\mathrm{E}(\mathbf{n}_t\mathbf{n}_t^H)=\sigma^2\mathbf{I}_n$. Under the assumption that $\mathbf{x}_t$ are independent identically zero-mean complex circular Gaussian distributed RVs, and if no prior information about $\mathbf{A}$ is used, the MDL

estimator is based on the eigenvalues of $\widehat{\mathbf{R}}$ and is given by the following minimizer [8].

$$\widehat{r} = \mathrm{Arg}\left\{\min_k \Lambda_k\right\} \quad \text{with} \quad \Lambda_k = T(n-k)\ln\left(\frac{\widehat{a}_k}{\widehat{g}_k}\right) + \frac{1}{2}k(2n-k)\ln T, \tag{12}$$

with $\widehat{a}_k \overset{\text{def}}{=} \frac{1}{n-k}\sum_{i=k+1}^{n}\widehat{\lambda}_i$ and $\widehat{g}_k \overset{\text{def}}{=} \prod_{i=k+1}^{n}\widehat{\lambda}_i^{1/(n-k)}$. The events $\widehat{r} < r$ and $\widehat{r} > r$ are called underestimation and overestimation, respectively. Since $(\Lambda_k)_{k=0,..,n-1}$ are functions of the eigenvalues $(\widehat{\lambda}_i)_{i=1,..,n}$ of $\widehat{\mathbf{R}}$, the derivation of the probabilities $P(\widehat{r} > r)$ and $P(\widehat{r} < r)$ needs the joint exact or asymptotic distribution of $(\lambda_i)_{i=1,..,n}$. Unfortunately, these two distributions are only available for circular complex Gaussian distribution [1], and are furthermore too complicated to be useful for the statistical analysis of the estimator $\widehat{r}$. Therefore, for simplifying the derivation of these probabilities, it has been argued [10], [11] and [22] by extended Monte Carlo experiments (essentially for $r = 1$ and $r = 2$) that

$$P(\widehat{r} > r) \approx P(\widehat{r} = r+1) \approx P(\Lambda_{r+1} < \Lambda_r) \quad \text{and} \quad P(\widehat{r} < r) \approx P(\widehat{r} = r-1) \approx P(\Lambda_{r-1} < \Lambda_r).$$

As the probability of overestimation is concerned, exact and approximate asymptotic upper bound of this probability has been derived in [22] showing that generally $P(\widehat{r} > r) \ll 1$. Therefore, we concentrate on the probability of underestimation for which (12) gives straightforwardly (see [12] for $r = 1$ and 2)

$$P(\Lambda_{r-1} < \Lambda_r) = P\left(H_r\left(\frac{\widehat{\lambda}_r}{\widehat{a}_r}\right) < T_r\right) \tag{13}$$

where $H_r(x) \overset{\text{def}}{=} \ln\left(\frac{1}{x}\left(1 + \frac{x-1}{n-r+1}\right)^{n-r+1}\right)$ and $T_r \overset{\text{def}}{=} \frac{1}{2T}(2n-2r+1)\ln T$. Because $H_r(x)$ is an increasing function for $x > 1$ with $H_r(1) = 0$ and $T_r > 0$, (13) is given by

$$P(\Lambda_{r-1} < \Lambda_r) = P\left(\frac{\widehat{\lambda}_r}{\widehat{a}_r} < T'_r\right), \tag{14}$$

where $T'_r = H_r^{-1}(T_r)$ is the unique solution of the equation $H_r(x) = T_r$ for $x > 1$.

To proceed, we must know the distribution of the ratio $\widehat{\lambda}_r/\widehat{a}_r$, which, derived from the exact or asymptotic distribution of $(\widehat{\lambda}_i)_{i=1,..,n}$, is also too complicated to give some insight. So we must resort to the approximation, used by all authors that have tackled this point, that the standard deviation of $\widehat{a}_r$ can be considered as negligible with respect to $\mathrm{E}(\widehat{a}_r)$. Consequently $\widehat{a}_r \approx \mathrm{E}(\widehat{a}_r)$, where [12] has refined the approximation $\mathrm{E}(\widehat{a}_r) \approx \sigma^2$ used in [10],[11] by taking into account the bias of the estimates $(\widehat{\lambda}_i)_{i=1,..,r}$. Using $\mathrm{E}(\widehat{a}_r) = \sigma^2 + \frac{1}{n-r}\sum_{i=1}^{r}\left(\lambda_i - \mathrm{E}(\widehat{\lambda}_i)\right)$ and (5), we have $\widehat{a}_r \approx m_r$ with

$$m_r \overset{\text{def}}{=} \sigma^2 - \frac{1}{T(n-r)}\sum_{i=1}^{r}\sum_{1\le j\neq i\le n}\frac{\lambda_i\lambda_j + |\lambda_{i,j}|^2 + \lambda_{i,j,i,j}}{\lambda_i - \lambda_j}. \tag{15}$$

Consequently $\widehat{\lambda}_r/\widehat{a}_r$ is approximately asymptotically Gaussian distributed with mean $\mu_r = \mathrm{E}(\widehat{\lambda}_r)/m_r$ and

variance $\sigma_r^2 = \mathrm{Var}(\widehat{\lambda}_r)/m_r^2$ where from (5) and (6)

$$\mu_r = \frac{1}{m_r}\left(\lambda_r + \frac{1}{T}\sum_{1\le j\ne r\le n}\frac{\lambda_r\lambda_j + |\lambda_{r,j}|^2 + \lambda_{r,j,r,j}}{\lambda_r - \lambda_j}\right) \quad \text{and} \quad \sigma_r^2 = \frac{1}{m_r^2 T}\left(\lambda_r^2 + |\lambda_{r,r}|^2 + \lambda_{r,r,r,r}\right) \tag{16}$$

and the probability of underestimation is approximately asymptotically given by

$$P(\widehat{r} < r) \approx 1 - Q\left(\frac{T_r' - \mu_r}{\sigma_r}\right), \tag{17}$$

with $Q(x) \stackrel{\text{def}}{=} \int_x^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}} dt$.

## IV. DOA ILLUSTRATIONS

To illustrate these general results, the detection of the number $r = 1$ or $r = 2$ of noncircular or/and nonGaussian sources impinging on an array of $n$ sensors is now considered. The common model for the received signal $\mathbf{x}_t$ is given by

$$\mathbf{x}_t = \mathbf{A}(\mathbf{\Theta})\mathbf{s}_t + \mathbf{n}_t$$

where $\mathbf{A}(\mathbf{\Theta}) = (\mathbf{a}_1, ..., \mathbf{a}_r)$ is the full column rank steering matrix where each vector $\mathbf{a}_k$ denotes the steering vector of the $k$-th source of DOA $\theta_k$. $\mathbf{n}_t$ is assumed circular complex Gaussian distributed, but $(s_{t,k})_{k=1,...,r}$ that have finite fourth-order moments, are not necessarily circular complex Gaussian distributed. $\mathrm{E}|s_{t,k}^2| = \sigma_{s_k}^2$, $\mathrm{E}(s_{t,k}^2) = \sigma_{s_k}^2 \rho_{s_k} e^{2\iota\phi_{s_k}}$ and $\mathrm{Cum}(s_{t,k}, s_{t,k}^*, s_{t,k}, s_{t,k}^*) = \sigma_{s_k}^4 \kappa_{s_k}$ where $\rho_{s_k} \in [0,1]$, $\phi_{s_k} \in [0,\pi)$ and $\kappa_{s_k} \in [-2, +\infty)$ denote the noncircularity rate and phase, and the kurtosis of $s_{t,k}$, respectively.

In all the numerical illustrations and Monte Carlo experiments, a uniform linear array of omni-directional $n = 5$ sensors, and half-wavelength spacing is used. Its centroid at the origin of the phase is used[6] (i.e., $\mathbf{a}_k = (e^{-\iota 2\pi \sin\theta_k}, e^{-\iota\pi\sin\theta_k}, 1, e^{\iota\pi\sin\theta_k}, e^{\iota 2\pi\sin\theta_k})^T$ where $\theta_k$ is the angle of DOA away from the broadside of the array). For each point on the figures, $T = 200$ snapshots (except in Fig.1) are taken and 10000 Monte Carlo runs are carried out to estimate the probabilities of underestimating.

We first consider the robustness of the MDL criterion to the distribution of the sources. For a single source ($r = 1$), it is straightforward to derive from $\mathbf{R}_x = \sigma_{s_1}^2 \mathbf{a}_1\mathbf{a}_1^H + \sigma^2\mathbf{I}_n$, $\mathbf{R}_x' = \sigma_{s_1}^2\rho_{s_1}e^{2\iota\phi_{s_1}}\mathbf{a}_1\mathbf{a}_1^T$, $\mathbf{Q}_x = \sigma_{s_1}^4\kappa_{s_1}(\mathbf{a}_1^* \otimes \mathbf{a}_1)(\mathbf{a}_1^T \otimes \mathbf{a}_1^H)$, the values $\lambda_1 = \|\mathbf{a}_1\|^2\sigma_{s_1}^2 + \sigma^2$, $\lambda_{1,j} = \delta_{1,j}\|\mathbf{a}_1\|^2\sigma_{s_1}^2\rho_{s_1}e^{2i\phi_{s_1}}$ and $\lambda_{1,j,1,j} = \delta_{1,j}\|\mathbf{a}_1\|^4\sigma_{s_1}^4\kappa_{s_1}$ that give from (15),(16), the expressions

$$m_1 = \sigma^2\left(1 - \frac{1}{T}\left(1 + \frac{\sigma^2}{\|\mathbf{a}_1\|^2\sigma_{s_1}^2}\right)\right) \tag{18}$$

$$\mu_1 = \frac{\sigma_{s_1}^2}{m_1}\left(\|\mathbf{a}_1\|^2 + \frac{\sigma^2}{\sigma_{s_1}^2}\right)\left(1 + \frac{n-1}{T}\frac{\sigma^2}{\|\mathbf{a}_1\|^2\sigma_{s_1}^2}\right) \tag{19}$$

[6]For rectilinear sources, the DOA $\theta_k$ and phase $\phi_k$ of noncircularity parameters are coupled [15] and as the performance depends on $\Delta\phi = \phi_2 - \phi_1$, the centroid of the array must be specified for fixing the performance.

$$\sigma_1^2 = \frac{\|\mathbf{a}_1\|^4 \sigma_{s_1}^4}{T m_1^2} \left[ \left( 1 + \frac{\sigma^2}{\|\mathbf{a}_1\|^2 \sigma_{s_1}^2} \right)^2 + \rho_{s_1}^2 + \kappa_{s_1} \right], \tag{20}$$

that specify the probability of underestimation $P(\widehat{r} = 0/r = 1)$ (17). Consequently this probability of underestimation has the following behavior: $P(\widehat{r} = 0/r = 1) = 1/2$ for the value $\text{SNR}_{1/2}$ of the signal-to-noise-ratio $\sigma_{s_1}^2/\sigma^2$ solution of $T_1' = \mu_1$ that does not depend on the distribution of the source. With respect to the circular Gaussian distribution, the probability of underestimation is larger ([resp. smaller]) for SNR $>$ $\text{SNR}_{1/2}$ ([resp. for SNR $<$ $\text{SNR}_{1/2}$]) for noncircular or/and nonGaussian distributions such that $\rho_{s_1}^2 + \kappa_{s_1} > 0$. The opposite behavior happens for distributions such that $\rho_{s_1}^2 + \kappa_{s_1} < 0$.

This behavior is illustrated in Fig.1 for the following three distributions: circular Gaussian ($\rho_{s_1} = \kappa_{s_1} = 0$), binary phase shift keying (BPSK) ($\rho_{s_1} = 1$ and $\kappa_{s_1} = -2$) and impulsive that takes the values $\{-1, 0, +1\}$ with $P(s_{t,1} = -1) = P(s_{t,1} = +1) = 1/2p$ and $P(s_{t,1} = 0) = 1 - 1/p$ for which $\rho_{s_1} = 1$ and $\kappa_{s_1} = p - 3$. We see from this figure that the probability of underestimation is sensitive to the distribution of the source, particularly for sources ed by (20).

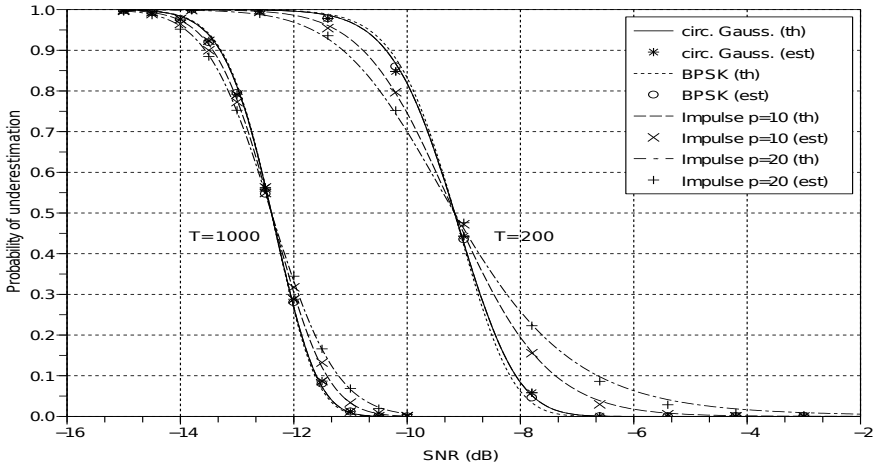


Fig.1 $P(\widehat{r} = 0/r = 1)$ as a function of the SNR for four distributions of the source and two values of the number $T$ of snapshots.

For two sources, the expressions of $m_2$, $\mu_2$ and $\sigma_2^2$ can be derived as well, showing their dependence on physical parameters such as number of observed snapshots, number of sensors, signal and noise powers, correlations, noncircularity rates, kurtosis and angular separation of the sources (see e.g., the expressions of the eigenvalues of $\mathbf{R}_x$ for the circular Gaussian distributions in [23, Section 5.4]). Such expressions are too complicated to analyse, but their numerical computations show that the probability of underestimation $P(\widehat{r} < 2/r = 2)$ is sensitive to the distribution of the sources particularly for large kurtosis.

**Remark 7:** We note that this property of sensitivity of the probability of underestimation given by the MDL criterion contrasts with the performance in terms of variance [14] and resolving power [15] derived from subspace algorithms issued from $\mathbf{R}_x$ that are robust to noncircular or/and nonGaussian distributions of the sources.

For possibly noncircular signals $\mathbf{x}_t$, it is well known that the DOA estimation may be improved in terms of accuracy [14] and resolving power [15] if the standard covariance matrix $\mathbf{R}_x = \mathrm{E}(\mathbf{x}_t \mathbf{x}_t^H)$ is replaced by the

augmented covariance matrix $\mathbf{R}_{\tilde{x}} = \mathrm{E}(\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t^H)$ with $\tilde{\mathbf{x}}_t \overset{\text{def}}{=} (\mathbf{x}_t^T, \mathbf{x}_t^H)^T$. More precisely, it is proved in [14],[15] that these performance are drastically improved only for rectilinear uncorrelated sources for which[7]

$$\mathbf{R}_x = \sum_{k=1}^{r} \sigma_{s_k}^2 \mathbf{a}_k\mathbf{a}_k^H + \sigma^2\mathbf{I}_n \quad \text{and} \quad \mathbf{R}_{\tilde{x}} = \sum_{k=1}^{r} \sigma_{s_k}^2 \tilde{\mathbf{a}}_k\tilde{\mathbf{a}}_k^H + \sigma^2\mathbf{I}_{2n} \tag{21}$$

with $\tilde{\mathbf{a}}_k \overset{\text{def}}{=} (\mathbf{a}_k^T, e^{-2\imath\phi_k}\mathbf{a}_k^H)^T$. Noting that both Result 1 and the approximate probability of underestimation (17) are also valid for $\mathbf{R}_{\tilde{x}}$ which is the covariance of $\tilde{\mathbf{x}}_t$, we can compare the probability of underestimation given by the MDL criterion associated with $\mathbf{R}_x$ and $\mathbf{R}_{\tilde{x}}$ whose eigenvalues are denoted $(\tilde{\lambda}_i)_{i=1,\ldots,2n}$. For a single rectilinear source the gap between $\tilde{\lambda}_1$ and $\sigma^2$ increases because now $\tilde{\lambda}_1 = 2\|\mathbf{a}_1\|^2\sigma_{s_1}^2 + \sigma^2$ against to $\lambda_1 = \|\mathbf{a}_1\|^2\sigma_{s_1}^2 + \sigma^2$ and thus the detection performance ought to improve. This is proved by the comparison of the new expressions of $m_1$, $\mu_1$ and $\sigma_1^2$ deduced from

$$\mathbf{R}'_{\tilde{x}} \overset{\text{def}}{=} \mathrm{E}(\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t^T) = \sigma_{s_1}^2\tilde{\mathbf{a}}_1\tilde{\mathbf{a}}_1^T + \sigma^2\begin{pmatrix} \mathbf{O} & \mathbf{I}_n \\ \mathbf{I}_n & \mathbf{O} \end{pmatrix} \quad \text{and} \quad \mathbf{Q}_{\tilde{x}} = \sigma_{s_1}^4\kappa_{s_1}(\tilde{\mathbf{a}}_1^* \otimes \tilde{\mathbf{a}}_1)(\tilde{\mathbf{a}}_1^T \otimes \tilde{\mathbf{a}}_1^H)$$

where $(\mathbf{Q}_{\tilde{x}})_{i+(j-1)2n,l+(k-1)2n} \overset{\text{def}}{=} \mathrm{Cum}((\tilde{\mathbf{x}}_t)_i, (\tilde{\mathbf{x}}_t)_j^*, (\tilde{\mathbf{x}}_t)_k, (\tilde{\mathbf{x}}_t)_l^*)$, that are given by (18), (19) and (20) where $\|\mathbf{a}_1\|^2$ is replaced by $2\|\mathbf{a}_1\|^2$. We obtain similar behaviors of the probability of underestimation that for the MDL criterion associated with $\mathbf{R}_x$, but the performance is improved as it is shown in Fig.2 for which the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ outperforms this criterion associated with $\mathbf{R}_x$ by about 1.5dB.

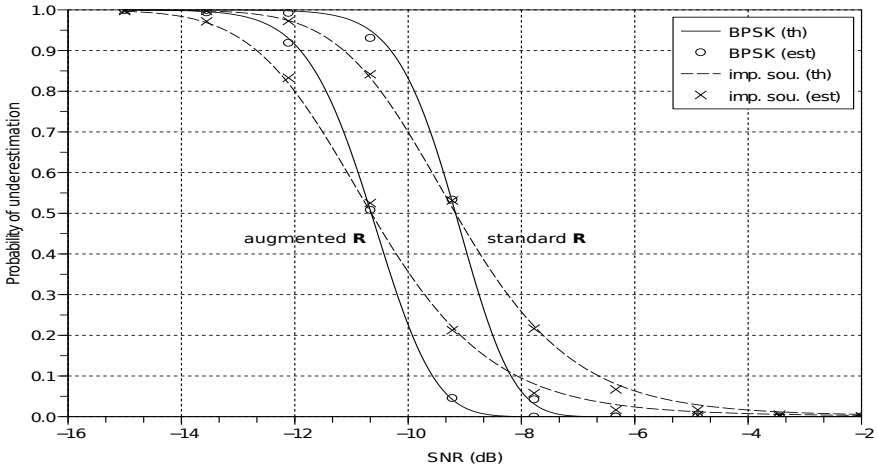


Fig.2 $P(\widehat{r} = 0/r = 1)$ given by the MDL criterion associated with $\mathbf{R}_x$ and $\mathbf{R}_{\tilde{x}}$ as a function of the SNR for BPSK and impulsive ($p = 20$) distributions of the source.

For two sources, the rank of $\mathbf{R}_{\tilde{x}} - \sigma^2\mathbf{I}_{2n}$ of $\tilde{\mathbf{x}}_t$ is $r = 4$ except for singularity cases. In particular for uncorrelated rectilinear sources this rank is 2. Intensive numerical computations and Monte Carlo experiments of $P(\widehat{r} < 4/r = 4)$ given by the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ for two correlated or/and nonrectilinear sources show that this MDL criterion is largely outperformed by the MDL criterion associated with $\mathbf{R}_x$. This is

[7]We note that if the sources are nonrectilinear or/and correlated, the number of components (i.e., the rank of $\mathbf{R}_{\tilde{x}} - \sigma^2\mathbf{I}_{2n}$) of $\tilde{\mathbf{x}}_t$ is generally $2r$.

explained by the smaller eigenvalue spread of the free noise augmented covariance matrix w.r.t. the covariance matrix. This implies that the gap between $\tilde{\lambda}_4$ and $\sigma^2$ is smaller than the one between $\lambda_2$ and $\sigma^2$. Consequently we concentrate now on two uncorrelated rectilinear sources.

Fig.3 and Fig.4 show the probability of underestimation given by the MDL criterion for two equipowered BPSK uncorrelated rectilinear sources where the SNR is defined by the ratio $\sigma_{s_1}^2/\sigma^2$. The probabilities of underestimation given by the MDL criterion associated with $\mathbf{R}_x$ and $\mathbf{R}_{\tilde{x}}$ are compared in Fig.3 as a function of the SNR for two DOA separations. We see that the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ largely outperforms those based on $\mathbf{R}_x$ by about 1.5dB.

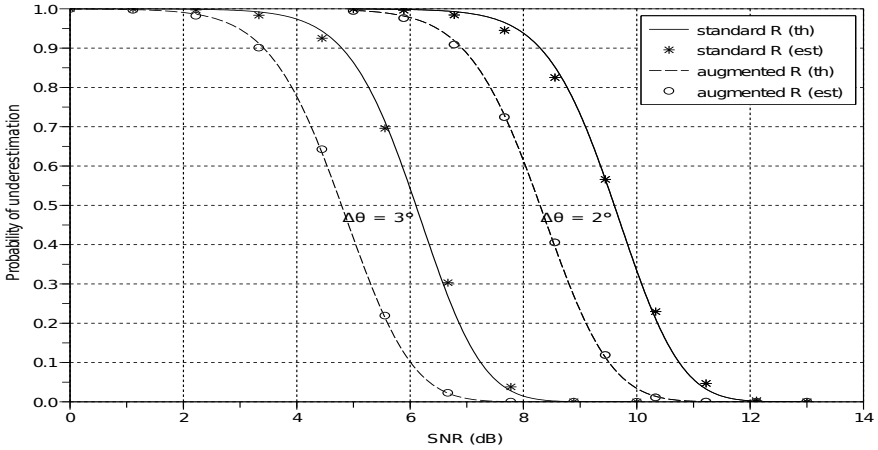


Fig.3 $P(\widehat{r} < 2/r = 2)$ given by the MDL criterion associated with $\mathbf{R}_x$ and $\mathbf{R}_{\tilde{x}}$ as a function of the SNR for two DOA separations $\Delta\theta$ (with $\theta_1 = -\theta_2 = \Delta\theta/2$) for two uncorrelated BPSK sources.

Fig.4 shows the probability of underestimation of the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ for two sources as a function of $\Delta\phi \stackrel{\text{def}}{=} \phi_2 - \phi_1$ for three SNR. We see that $\Delta\phi = \pi/2$ optimizes the capability of detection. This property is similar with the performance in terms of variance [14] and resolving power [15] that are likewise optimized for $\Delta\phi = \pi/2$.

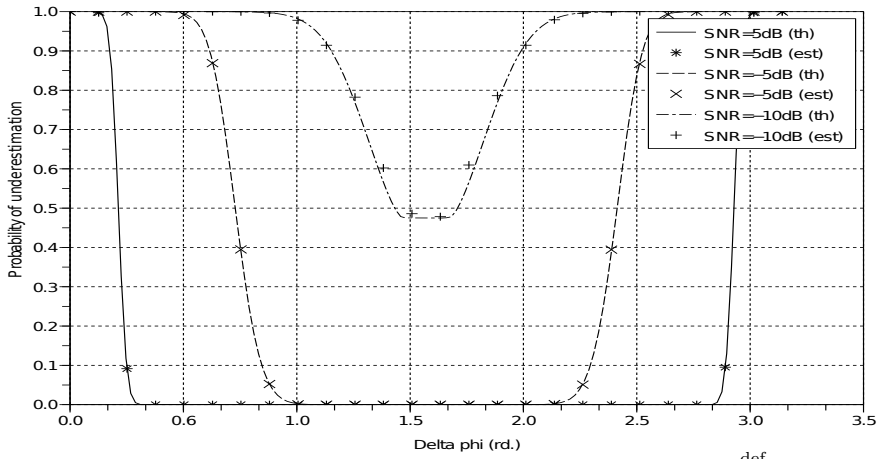


Fig.4 $P(\widehat{r} < 2/r = 2)$ given by the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ as a function of $\Delta\phi \stackrel{\text{def}}{=} \phi_1 - \phi_2$ for two uncorrelated BPSK sources with $\Delta\theta = 1$deg for three SNR.

Fig.5 illustrates that the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ can also detect two uncorrelated rectilinear

sources of common DOA but with different phases of noncircularity. We see that naturally the probability of underestimating decreases when $\Delta\phi$ increases.

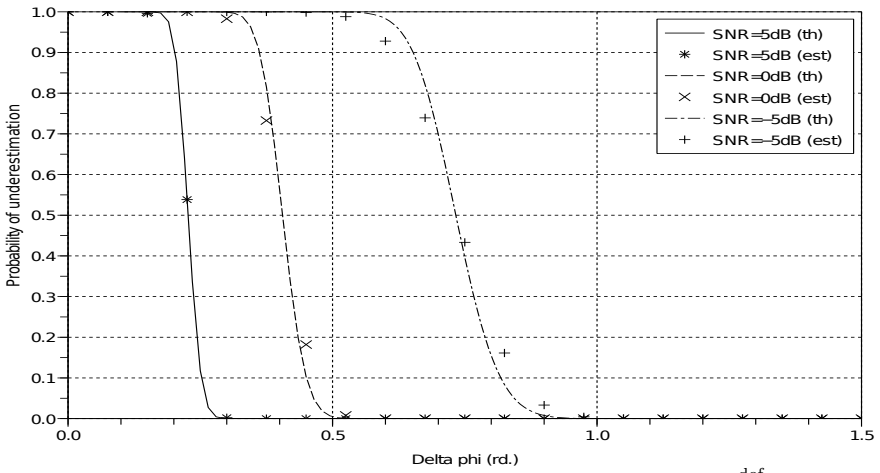


Fig.5 $P(\hat{r} < 2/r = 2)$ given by the MDL criterion associated with $\mathbf{R}_{\tilde{x}}$ as a function of $\Delta\phi \stackrel{\text{def}}{=} \phi_1 - \phi_2$ for two uncorrelated BPSK sources of same DOA for three SNR.

Finally we note that our asymptotic theoretical analysis allows us to perfectly predict the threshold region in all the scenarios we have considered. Furthermore, the different probabilities of underestimating estimated by Monte Carlo experiments, fit the asymptotic theoretical ones for relatively small number of snapshots $T = 200$, except for the particular scenarios for which some of the eigenvalues $(\tilde{\lambda}_i)_{i=1,..r}$ are very closed or $\tilde{\lambda}_r$ is very closed to $\sigma^2$. For example, in the case of two equipowered uncorrelated rectilinear sources impinging on a uniform linear array, the eigenvalues $(\tilde{\lambda}_i)_{i=1,2}$ are given by

$$\tilde{\lambda}_i = 2\|\mathbf{a}_1\|^2\sigma_{s_1}^2\left(1 + (-1)^i\cos((M-1)\frac{\Delta\theta}{2} - \Delta\phi)\frac{\sin(M\frac{\Delta\theta}{2})}{M\sin(\frac{\Delta\theta}{2})}\right) + \sigma^2 \quad i = 1, 2.$$

This happens in particular for $\Delta\theta \approx 0$ in two cases : $\Delta\phi$ approaches $\pi/2$ for which $\tilde{\lambda}_1 \approx \tilde{\lambda}_2$ and very weak $\Delta\phi$ for which $\tilde{\mathbf{a}}_1$ and $\tilde{\mathbf{a}}_2$ get very closed and $\tilde{\lambda}_2 \approx \sigma^2$. In these two cases the asymptotic theoretical value of $\mathrm{E}(\widehat{\tilde{\lambda}}_2)$ derived from (5) is no longer valid because $T = 200$ or equivalently the SNR is not large enough. The following table gives the minimum and maximun value of $\Delta\phi$ for which our asymptotic approximation is valid for $\Delta\theta = 0$. We clearly see from this table that the domain of validity of our asymptotic approximation enlarges when $T$ or the SNR increases.

| $T$ | 200 | | $10^4$ | | $10^5$ | |
|---|---|---|---|---|---|---|
| $\Delta\phi$ | min | max | min | max | min | max |
| SNR= $-5$dB | 0.30rd | 1.50rd | 0.07rd | 1.56rd | 0.04rd | 1.56rd |
| SNR= 0dB | 0.15rd | 1.54rd | 0.05rd | 1.56rd | 0.02rd | 1.56rd |
| SNR= $+5$dB | 0.07rd | 1.54rd | 0.02rd | 1.56rd | 0.01rd | 1.56rd |

Table 1 $\Delta\phi_{\min}$ and $\Delta\phi_{\max}$ for two uncorrelated BPSK sources with $\Delta\theta = 0$deg.

## APPENDIX

### Proof of Lemma 1

From the independence of the samples $(\mathbf{x}_t)_{t=1,..,T}$, we straightforwardly have

$$\mathrm{E}[(\mathbf{a}_1^H\delta\mathbf{R}\mathbf{a}_2)(\mathbf{a}_3^H\delta\mathbf{R}\mathbf{a}_4)] = \frac{1}{T}\left\{\mathrm{E}[(\mathbf{a}_1^H\mathbf{x}_t)(\mathbf{x}_t^H\mathbf{a}_2)(\mathbf{a}_3^H\mathbf{x}_t)(\mathbf{x}_t^H\mathbf{a}_4)] - (\mathbf{a}_1^H\mathbf{R}\mathbf{a}_2)(\mathbf{a}_3^H\mathbf{R}\mathbf{a}_4)\right\}. \tag{22}$$

Applying the following identity for arbitrary zero-mean RVs $(z_k)_{k=1,..,4}$

$$\mathrm{E}(z_1z_2z_3z_4) = \mathrm{E}(z_1z_2)\mathrm{E}(z_3z_4) + \mathrm{E}(z_1z_3)\mathrm{E}(z_2z_4) + \mathrm{E}(z_1z_4)\mathrm{E}(z_2z_3) + \mathrm{Cum}(z_1,z_2,z_3,z_4)$$

to the expectation of the right hand side of (22), then using the multilinearity of the cumulant

$$\begin{aligned}\mathrm{Cum}(\mathbf{a}_1^H\mathbf{x}_t, \mathbf{x}_t^H\mathbf{a}_2, \mathbf{a}_3^H\mathbf{x}_t, \mathbf{x}_t^H\mathbf{a}_4) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_{1,i}^* a_{2,j} a_{3,k}^* a_{4,l}\mathrm{Cum}(x_{t,i}, x_{t,j}^*, x_{t,k}, x_{t,l}^*)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(\mathbf{a}_2^T\otimes\mathbf{a}_1^H)_{i+(j-1)n}(\mathbf{Q})_{i+(j-1)n,l+(k-1)n}(\mathbf{a}_3^*\otimes\mathbf{a}_4)_{l+(k-1)n},\end{aligned}$$

Lemma 1 is proved. ■

**Proof of Result 1**

We follow here the Wilkinson's approach [3, p.68], carried on by Kaveh *et al* [5], that we restate for the benefit of the reader. Define $\widehat{\mathbf{R}} = \mathbf{R} + \delta\mathbf{R}$ in terms of a random perturbation $\mathbf{\Delta}$ to $\mathbf{R}$ with a perturbation factor $\epsilon$. Thus $\delta\mathbf{R} = \epsilon\mathbf{\Delta}$, where $\mathbf{\Delta}$ characterizes the direction of the zero-mean random perturbation term $\delta\mathbf{R}$.

Let $\bar{\mathbf{v}}_i$ denote the unnormalized eigenvector of $\widehat{\mathbf{R}}$ specified by (2), (3). It is given by a perturbation expansion [3, p.68]

$$\bar{\mathbf{v}}_i = \mathbf{v}_i + \sum_{1\le j\ne i\le n}\left(\sum_{k=1}^{\infty} t_{k,j}^{(i)}\epsilon^k\right)\mathbf{v}_j, \tag{23}$$

which gives to the second order

$$\widehat{\mathbf{v}}_i = \frac{\bar{\mathbf{v}}_i}{||\bar{\mathbf{v}}_i||} = \left(1 - \frac{1}{2}\sum_{1\le j\ne i\le n}|t_{1,j}^{(i)}|^2\epsilon^2\right)\mathbf{v}_i + \sum_{1\le j\ne i\le n}\left(t_{1,j}^{(i)}\epsilon + t_{2,j}^{(i)}\epsilon^2\right)\mathbf{v}_j + o(\epsilon^2). \tag{24}$$

Using $\widehat{\mathbf{R}}\bar{\mathbf{v}}_i = \widehat{\lambda}_i\bar{\mathbf{v}}_i$, $\mathbf{v}_i^H\mathbf{R}\mathbf{v}_j = \delta_{i,j}\lambda_i$ and (23), we obtain

$$\widehat{\lambda}_i = \lambda_i + \epsilon(\mathbf{v}_i^H\mathbf{\Delta}\mathbf{v}_i) + \epsilon^2\sum_{1\le j\ne i\le n} t_{1,j}^{(i)}(\mathbf{v}_i^H\mathbf{\Delta}\mathbf{v}_j) + o(\epsilon^2) \tag{25}$$

and for $k \ne i$,

$$t_{1,k}^{(i)} = \frac{\mathbf{v}_k^H\mathbf{\Delta}\mathbf{v}_i}{\lambda_i - \lambda_k} \quad \text{and} \quad t_{2,k}^{(i)} = -\frac{(\mathbf{v}_i^H\mathbf{\Delta}\mathbf{v}_i)(\mathbf{v}_k^H\mathbf{\Delta}\mathbf{v}_i)}{(\lambda_i-\lambda_k)^2} + \sum_{1\le j\ne i\le n}\frac{(\mathbf{v}_k^H\mathbf{\Delta}\mathbf{v}_j)(\mathbf{v}_j^H\mathbf{\Delta}\mathbf{v}_i)}{(\lambda_i-\lambda_j)(\lambda_i-\lambda_k)}.$$

Because $\mathrm{E}(\widehat{\mathbf{R}}) = \mathbf{R}$, $\epsilon\mathrm{E}(t_{1,k}^{(i)}) = 0$ and using Lemma 1, we obtain (where we note that $\epsilon^2\mathrm{E}(t_{2,k}^{(i)}) = 0$ in [5])

for $k \neq i$ and $l \neq j$

$$\begin{aligned}
\epsilon^2 \mathrm{E}(t_{1,k}^{(i)} t_{1,l}^{(j)^*}) &= \frac{1}{T} \frac{(\lambda_i \lambda_k \delta_{i,j} \delta_{k,l} + \lambda_{j,k} \lambda_{i,l}^* + \lambda_{i,k,j,l})}{(\lambda_i - \lambda_k)(\lambda_j - \lambda_l)} \\
\epsilon^2 \mathrm{E}(t_{1,k}^{(i)} t_{1,l}^{(j)}) &= \frac{1}{T} \frac{(\lambda_i \lambda_j \delta_{j,k} \delta_{i,l} + \lambda_{k,l} \lambda_{i,j}^* + \lambda_{i,k,l,j})}{(\lambda_i - \lambda_k)(\lambda_j - \lambda_l)} \\
\epsilon^2 \mathrm{E}(t_{2,k}^{(i)}) &= -\frac{1}{T} \frac{(\lambda_{i,k} \lambda_{i,i}^* + \lambda_{i,i,k,i})}{(\lambda_i - \lambda_k)^2} + \frac{1}{T} \sum_{1 \leq j \neq i \leq n} \frac{(\lambda_{i,j} \lambda_{j,k}^* + \lambda_{j,k,j,i})}{(\lambda_i - \lambda_j)(\lambda_i - \lambda_k)}.
\end{aligned}$$

Plugging these four expectations in the expectations and covariances of $\widehat{\mathbf{v}}_i$ and $\widehat{\lambda}_i$ given in (24) and (25) respectively, straightforwardly proves Result 1. ■

## REFERENCES


[1] T.W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, pp. 122-148, 1963.

[2] G.P. Gupta, "Asymptotic theory for principal component analysis in the complex case," *J. Indian Stat. Assoc.*, vol. 3, pp. 97-106, 1965.

[3] J.H. Wilkinson, *The algebraic eigenvalue principles*, New York: Oxford University Press, 1965.

[4] D.R. Brillinger, *Time series, data analysis and theory*, Extended edition, Holden-Day, INC, 1980.

[5] M. Kaveh and A.J. Barabell, "The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise," *IEEE Trans. on ASSP*, vol. 34, no. 2, pp. 331-341, April 1986.

[6] B. Friedlander and A.J. Weiss, "On the second-order statistics of the eigenvectors of sample covariance matrices," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3136-3139, November 1998.

[7] J. Rissanen, "Modeling by shortest data description," *Artomatica*, vol. 14, pp. 465-471, 1978.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. on ASSP*, vol. 33, no. 2, pp. 387-392, April 1985.

[9] X.L. Li, M. Anderson, and T. Adalı, "Principal component analysis for noncircular signals in the presence of circular white Gaussian noise," *Proc. Asilomar Conference on Signals, Systems and Computers.*, 2010.

[10] H. Wang and M. Kaveh, "On the performance of signal subspace processing - Part I: narrow-band systems," *IEEE Trans. on ASSP*, vol. 34, no. 5, pp. 1201-1209, October 1986.

[11] M. Kaveh, H. Wang, and H. Hung, "On the theoretic performance of a class of estimators of the number of narrow-band sources," *IEEE Trans. on ASSP*, vol. 35, no. 9, pp. 1350-1352, September 1987.

[12] F. Haddadi, M.M. Mohammadi, M.M. Nayebi, and M.R. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452-457, May 2010.

[13] F. Nesser and J. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293-1302, July 1993.

[14] H. Abeida and J.P. Delmas, "MUSIC-like estimation of direction of arrival for non-circular sources," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2678-2690, July 2006.

[15] H. Abeida and J.P. Delmas, "Statistical performance of MUSIC-like algorithms in resolving noncircular sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4317-4329, September 2008.

[16] P.J. Schreier and L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714-725, March 2003.

[17] J.R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*, Wiley series in probability and statistics, revised edition, 1999.

[18] J.P. Delmas and H. Abeida, "Asymptotic distribution of circularity coefficients estimate of complex random variables," *Signal Processing (Elsevier)*, vol. 89, pp. 2311-2698, December 2009.

[19] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, 1980.

[20] D.R. Cox and D.V. Hinkley, *Theoretical statistics*, Chapman and Hall, London, 1974.

[21] H. Abeida and J.P. Delmas, "Efficiency of subspace-based DOA estimators," *Signal Processing (Elsevier)*, vol. 87, no. 9, pp. 2075-2084, September 2007.

[22] W. Xu and M Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413-1426, June 1995.

[23] M. Kaveh and H. Wang, *Advances in Spectrum analysis and array processing, S. Haykin Editor*, Prentice Hall, 1991.