



HAL
open science

Looking for Equivalence between Maximum Likelihood and Sparse DOA Estimators

Thomas Aussaguès, Anne Ferréol, Alice Delmer, Pascal Larzabal

► **To cite this version:**

Thomas Aussaguès, Anne Ferréol, Alice Delmer, Pascal Larzabal. Looking for Equivalence between Maximum Likelihood and Sparse DOA Estimators. EUSIPCO 2024 - 32nd European Signal Processing Conference, EURASIP, Aug 2024, Lyon, France. hal-04666050v2

HAL Id: hal-04666050

<https://hal.science/hal-04666050v2>

Submitted on 20 Aug 2024 (v2), last revised 29 Aug 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Looking for Equivalence between Maximum Likelihood and Sparse DOA Estimators

Thomas Aussaguès^{*†}, Anne Ferréol^{*†}, Alice Delmer^{*} and Pascal Larzabal[†]

^{*}Thales, 4 avenue des Louvresses, 92230 Gennevilliers, France

[†]SATIE, Université Paris-Saclay, UMR CNRS 8029, 4 avenue des Sciences, 91190 Gif-sur-Yvette, France

Contact : thomas.aussagues@ens-paris-saclay.fr

Abstract—Sparse Direction-of-Arrival estimators depend on the regularization parameter λ which is often empirically tuned. In this work, conducted under the vectorized covariance matrix model, we are looking for theoretical equivalence between the Maximum Likelihood (ML) and sparse estimators. We show that under mild conditions, λ can be chosen thanks to the distribution of the minimum of the ML criterion in the case of two impinging sources. We derive this distribution under complex non-circular Gaussian noise. The corresponding λ choice is θ -invariant, only requiring an upper bound on the number of sources. Furthermore, it guarantees the global minimum of the sparse ℓ_0 -regularized criterion to be the ML solution.

Numerical experiments confirm that, for the proposed λ , sparse and ML estimators yield the same statistical performance.

Index Terms—DOA estimation, sparse estimation, regularization parameter, maximum likelihood

I. INTRODUCTION

Direction-of-Arrival (DOA) estimation is a classical signal processing problem for which numerous algorithms have been proposed through last decades [1]. However, conventional methods suffer from multiple limitations : Capon’s beamformer [2] can not separate close sources, MUSIC method [3] achieves super-resolution but can not handle highly correlated sources or work with few time samples. To overcome some of the foregoing issues, the signal processing community has widely investigated the topic of sparse signal representations [4].

In the field of DOA estimation, the concept of sparse representation has first been applied on the Single Measurement Vector (SMV) model. It aims to recover a sparse vector $\mathbf{x}_0 \in \mathbb{C}^G$ which has only M non-zero entries from a single noisy array snapshot $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n}$ where $\mathbf{A} \in \mathbb{C}^{N \times G}$, $G \gg N$ denotes an overcomplete basis and $\mathbf{n} \in \mathbb{C}^N$ a noise. The sparse estimation problem can be formulated as the minimization of the following non-convex constrained objective:

$$\min_{\mathbf{x} \in \mathbb{C}^G} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \beta \quad (1)$$

where $\|\cdot\|_0$ denotes the ℓ_0 -norm and $\beta > 0$ a constraint on the residuals. Several authors proposed to choose it [5]–[8] as an upper bound on the noise distribution.

Nonetheless, solving this NP-hard problem requires combinatorial optimization. Numerous greedy algorithms such as matching pursuit and orthogonal matching pursuit have been developed to solve (1). Aforementioned algorithms are based on iterative support reconstruction which makes them

extremely sensitive to support error which can not be undone. Some authors [6]–[9] suggested a convex- ℓ_1 relaxation of (1) to find a good approximate solution since ℓ_0 and ℓ_1 -formulations are equivalent under the RIP (Restricted Isometry Property). However, the RIP condition requires low correlations between vectors of \mathbf{A} which is rarely verified. Thus, ℓ_0 -regularized formulation of (1) is employed:

$$\min_{\mathbf{x} \in \mathbb{C}^G} \left\{ J_{\ell_0}(\lambda, \mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_0 \right\} \quad (2)$$

where the objective J_{ℓ_0} depends on the regularization parameter $\lambda > 0$ which balances the reconstruction error against the sparsity of the solution. Small values of λ lead to non-sparse solutions whereas larger ones enforce sparsity. Consequently, the choice of the regularization parameter is of paramount importance.

In this work, we look for theoretical equivalence between ML and sparse estimators. We propose to select the regularization parameter thanks to the distribution of the minimum of the ML criterion which we derive under complex non-circular Gaussian noise. This yields a θ -invariant choice for λ only requiring an upper bound on the number of sources. Numerical experiments show that with the proposed choice for λ , ML and sparse estimators have equal performance.

II. MAXIMUM LIKELIHOOD ON THE VECTORIZED COVARIANCE MATRIX MODEL

A. Vectorized covariance matrix model

We consider a scenario of M independent sources of directions $\Theta = \{\theta_1, \dots, \theta_M\}$ impinging on an array of N sensors. Assuming narrowband hypothesis, the array observation is given by:

$$\mathbf{x}(t) = \sum_{m=1}^M \mathbf{a}(\theta_m) s_m(t) + \mathbf{n}(t) = \mathbf{A}(\Theta) \mathbf{s}_\Theta(t) + \mathbf{n}(t) \quad (3)$$

with $\mathbf{A}(\Theta)$ the array manifold matrix formed by the steering vectors $\mathbf{a}(\theta_m) \in \mathbb{C}^N$, $m = 1 \dots M$, $\mathbf{s}_\Theta(t) = [s_1(t), \dots, s_M(t)]^T \in \mathbb{C}^M$ a vector containing the complex envelopes of the emitted signals and $\mathbf{n}(t) \in \mathbb{C}^N$ a complex Gaussian circular noise, independent of $\mathbf{s}(t)$, with covariance $\mathbb{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma^2 \mathbf{I}_N$ where $\mathbb{E}[\cdot]$ is the temporal mean, $(\cdot)^H$ denotes the complex conjugate transpose and \mathbf{I}_N the identity matrix of size N .

The array covariance matrix is then given by:

$$\mathbf{R}_x = \mathbb{E} [\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{A}(\boldsymbol{\Theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\Theta}) + \sigma^2\mathbf{I}_N \quad (4)$$

where $\mathbf{R}_s = \mathbb{E} [\mathbf{s}_\Theta(t)\mathbf{s}_\Theta^H(t)]$ is the sources covariance matrix.

In practice, the covariance matrix is not available. Instead, \mathbf{R}_x is replaced by its ML estimate $\widehat{\mathbf{R}}_x$ computed using K identically and independently distributed (i.i.d) array snapshots. Assuming temporally white noise ($\forall i \neq j, \mathbb{E} [\mathbf{n}^H(t_i)\mathbf{n}(t_j)] = 0$), $\widehat{\mathbf{R}}_x$ can be written as:

$$\widehat{\mathbf{R}}_x = \frac{1}{K} \sum_{k=1}^K \mathbf{x}(t_k)\mathbf{x}^H(t_k) = \mathbf{R}_x + \Delta\mathbf{R}_x \quad (5)$$

where $K\Delta\mathbf{R}_x$ is a Wishart noise due to the finite number of snapshots [10] since noise samples are independent from snapshot to snapshot.

Applying the previous model (3) for all K snapshots leads to a Multi Measurement Vectors (MMV) model which improves accuracy at the cost of computational efficiency [5]. Thanks to vectorization, the MMV model can be casted into an SMV model: the vectorized covariance matrix model [6], [8], [9], [11]. This model exploits all the K array snapshots while maintaining a low complexity. Furthermore, it relies on a virtual array of at most $N^2 - N$ non-redundant antennas which has a reduced beamwidth and lower sidelobes compared to the standard array [12].

Assuming that sources are uncorrelated, applying the vectorization operator $\text{vec}(\cdot)$ to $\widehat{\mathbf{R}}_x - \sigma^2\mathbf{I}_N$ yields:

$$\mathbf{r} = \text{vec}(\widehat{\mathbf{R}}_x - \sigma^2\mathbf{I}_N) = \sum_{m=1}^M \mathbf{b}(\theta_m)\gamma_m + \boldsymbol{\delta} = \mathbf{B}(\boldsymbol{\Theta})\boldsymbol{\gamma}_\Theta + \boldsymbol{\delta} \quad (6)$$

where $\mathbf{b}(\theta_m) = \mathbf{a}^*(\theta_m) \otimes \mathbf{a}(\theta_m)$ is the virtual array steering vector with \otimes the Kronecker product and $(\cdot)^*$ the complex conjugate. The corresponding mixing matrix is referred as $\mathbf{B}(\boldsymbol{\Theta}) = [\mathbf{b}(\theta_1), \dots, \mathbf{b}(\theta_M)]$. $\gamma_m = \mathbb{E} [|s_m(t)|^2]$ denotes the power of the m -th source and $\boldsymbol{\gamma}_\Theta = [\gamma_1, \dots, \gamma_M]^T$ the vector of the sources powers. Finally, $\boldsymbol{\delta} = \text{vec}(\Delta\mathbf{R}_x)$ is the vectorized noise.

According to [13], the vectorized observation asymptotically (for K sufficiently large) follows a multivariate complex Gaussian law $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$ with:

$$\begin{aligned} \boldsymbol{\mu} &= \mathbb{E} [\mathbf{r}] = \mathbf{B}(\boldsymbol{\Theta})\boldsymbol{\gamma}_\Theta \\ \boldsymbol{\Gamma} &= \mathbb{E} [(\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})^H] = \frac{1}{K}\mathbf{R}_x^T \otimes \mathbf{R}_x \\ \mathbf{C} &= \mathbb{E} [(\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})^T] = \boldsymbol{\Gamma}\mathbf{K} \end{aligned} \quad (7)$$

where \mathbf{C} is the second moment at the second order of \mathbf{r} and $(\cdot)^T$ denotes the transposition operator. $\mathbf{K} \in \mathbb{R}^{N^2 \times N^2}$ is the permutation matrix such that $\text{vec}(\mathbf{M}^T) = \mathbf{K}\text{vec}(\mathbf{M})$ for any square matrix \mathbf{M} of size N^2 .

B. Maximum Likelihood estimation

For additive Gaussian noise a ML estimator of $\boldsymbol{\Theta}$, known to achieve the Cramér-Rao lower bound at high Signal-to-Noise Ratio (SNR), can be easily obtained from (6) [14].

However, for non-diagonal covariance matrices such as (7) this yields a multi-term log-likelihood function. This objective can be simplified for diagonal covariance matrices leading us to apply a pre-whitening step to (6):

$$\mathbf{y} = \widehat{\mathbf{W}}\mathbf{r} = \widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\Theta})\boldsymbol{\gamma}_\Theta + \boldsymbol{\delta}_w \quad (8)$$

where $\widehat{\mathbf{W}} = \sqrt{K}\widehat{\mathbf{R}}_x^{-T/2} \otimes \widehat{\mathbf{R}}_x^{-1/2}$ is the estimated whitening matrix and $\boldsymbol{\delta}_w = \widehat{\mathbf{W}}\boldsymbol{\delta}$ a spatially white noise with $\mathbb{E} [\boldsymbol{\delta}_w\boldsymbol{\delta}_w^H] = \mathbf{I}_{N^2}$. Applying the ML to (8) yields the following estimator of $\boldsymbol{\Theta}$:

$$\begin{aligned} J_{ML}(\boldsymbol{\Theta}) &= \text{tr} (\boldsymbol{\Pi}^\perp(\boldsymbol{\Theta})\mathbf{y}\mathbf{y}^H) = \mathbf{y}^H\boldsymbol{\Pi}^\perp(\boldsymbol{\Theta})\mathbf{y} \\ \widehat{\boldsymbol{\Theta}} &= \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^M} J_{ML}(\boldsymbol{\Theta}) \end{aligned} \quad (9)$$

where $\boldsymbol{\Pi}^\perp(\boldsymbol{\Theta}) = \mathbf{I}_{N^2} - (\widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\Theta}))(\widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\Theta}))^\#$ is the noise projector computed for directions $\boldsymbol{\Theta}$ with $(\cdot)^\#$ the Moore-Penrose pseudo inverse and $\text{tr}(\cdot)$ the trace operator. In what follows, the optimal value of the ML criterion J_{ML} is referred as:

$$\epsilon = \min_{\boldsymbol{\Theta} \in \mathbb{R}^M} J_{ML}(\boldsymbol{\Theta}) = J_{ML}(\widehat{\boldsymbol{\Theta}}) = \mathbf{y}^H\boldsymbol{\Pi}^\perp(\widehat{\boldsymbol{\Theta}})\mathbf{y} \quad (10)$$

Nonetheless, the ML is restricted to small numbers of sources M in practice as it requires an M -dimensional optimization. In the next sections, we analyze the equivalence between ML and sparse estimators. We show that, for proper values of the regularization parameter, both methods leads to equal statistical performance.

III. SPARSE ESTIMATION

A. Sparse modeling

Under the hypothesis that DOAs lie within a predefined grid of directions $\boldsymbol{\Phi} = \{\varphi_1, \dots, \varphi_G\}$ ($\boldsymbol{\Theta} \subset \boldsymbol{\Phi}$), sparsity can be introduced in (8) using an overcomplete dictionary $\mathbf{B}(\boldsymbol{\Phi}) = [\mathbf{b}(\varphi_1), \dots, \mathbf{b}(\varphi_G)] \in \mathbb{C}^{N^2 \times G}$ with $G \gg N^2$. Thus, we obtain the following sparse model for the observation \mathbf{y} :

$$\mathbf{y} = \widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\Phi})\boldsymbol{\gamma}_0 + \boldsymbol{\delta}_w \quad (11)$$

where $\boldsymbol{\gamma}_0$ is a sparse vector which has exactly M non-zeros entries corresponding to the sources powers. Thereafter, $\boldsymbol{\gamma}_0$ is referred as the sparse spectrum.

B. Sparse problem formulation

The aim of DOA estimation is to find the directions of the impinging signals from the observation \mathbf{y} which corresponds, in the sparse model (11), to the non-zeros entries of $\boldsymbol{\gamma}_0$. Thus, we need to estimate the sparse spectrum $\boldsymbol{\gamma}_0$ in order to provide an estimate of $\boldsymbol{\Theta}$. This problem is ill-posed since $G \gg N^2$. Hence, sparsity of $\boldsymbol{\gamma}_0$ is exploited to ensure the uniqueness of the solution leading to the following constrained optimization problem:

$$\min_{\boldsymbol{\gamma} \in \mathbb{C}^G} \|\mathbf{y} - \widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\Phi})\boldsymbol{\gamma}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\gamma}\|_0 \leq M \quad (12)$$

Following Delmer's work [11], the solution of (12) can be obtained through the minimization of the ℓ_0 -regularized objective function given by:

$$\min_{\gamma \in \mathbb{C}^G} \left\{ J_{\ell_0}(\lambda, \gamma) = \frac{1}{2} \|\mathbf{y} - \widehat{\mathbf{W}}\mathbf{B}(\Phi)\gamma\|_2^2 + \lambda \|\gamma\|_0 \right\} \quad (13)$$

with a suitable regularization parameter $\lambda > 0$. This choice is discussed in section IV.

C. Use of an oracle grid

As mentioned in section III-A, DOA sparse estimates correspond to grid directions whose sparse spectrum components are non-zeros. Therefore, the sparse estimates are conditioned by the grid choice. Given that this paper focuses on the equivalence between ML and sparse methods, an oracle grid containing the ML estimates is employed.

IV. INFLUENCE OF THE REGULARIZATION PARAMETER

The choice of λ in (13) is of utmost importance since it controls the trade-off between estimation error and sparsity. Several authors [5], [7], [8], [15] proposed to fix λ using the noise level for ℓ_1 -norm. In [6], an optimal value for λ is derived using the Lagrangian. Additionally, one can use the L -curve [16] to choose λ . However, the dictionary $\widehat{\mathbf{W}}\mathbf{B}(\Phi)$ has high mutual coherence which makes the RIP condition unverified. Therefore, ℓ_1 -norm can not be employed and neither the previous techniques.

Regarding the ℓ_0 -framework, λ is generally empirically tuned given that few results are available. Recently, Delmer [11] *et al.* proposed to choose λ so that J_{ℓ_0} has a global minimum in $\hat{\gamma}$ ie.:

$$\lambda > 0 \quad \forall \gamma \in \mathbb{C}^G, J_{\ell_0}(\lambda, \hat{\gamma}) \leq J_{\ell_0}(\lambda, \gamma) \quad (14)$$

where $\hat{\gamma}$ is the sparse spectrum, in the neighboring of γ_0 , obtained using the ML DOA estimate $\hat{\Theta}$ (9). For a given observation \mathbf{y} , the values of λ that achieves condition (14) are within a stochastic admissible interval $I_M(\mathbf{y}) = [\lambda_M^-, \lambda_M^+]$. This interval ensures the global minimizer of (13) to be the ML solution which is an M -sparse vector.

For the case $M = 2$, the corresponding interval is $I_2(\mathbf{y}) = [\lambda_2^-, \lambda_2^+]$ where:

$$\lambda_2^- \approx \frac{c_2(\mathbf{y}) - c_3(\mathbf{y})}{2} \quad \text{and} \quad \lambda_2^+ \approx \frac{c_1(\mathbf{y}) - c_2(\mathbf{y})}{2} \quad (15)$$

with $c_k(\mathbf{y}) = \inf\{\|\mathbf{y} - \widehat{\mathbf{W}}\mathbf{B}(\Phi)\gamma\|_2^2, \|\gamma\|_0 = k\}$ the ML criterion optimal value for a prescribed sparsity level k . Note that we have $c_3(\mathbf{y}) \leq c_2(\mathbf{y}) \leq c_1(\mathbf{y})$. In the following, I_2 is assumed non-empty (*ie.* $c_1(\mathbf{y}) > c_3(\mathbf{y})$).

For arrays presenting robustness against second order ambiguities for $M = 2$ sources with equal power, the relation $c_2(\mathbf{y}) \ll c_1(\mathbf{y})$ holds. Thus, the following inequality can be easily proven:

$$\lambda_2^- \leq \frac{1}{2} c_2(\mathbf{y}) \leq \lambda_2^+ \quad (16)$$

A suitable choice for λ satisfying (16) is then:

$$\lambda = \frac{1}{2} c_2(\mathbf{y}) \quad (17)$$

To perform an off-line choice for λ , the distribution of $c_2(\mathbf{y})$ must be known. In the next section, we derive the distribution $c_2(\mathbf{y})$ which is also the minimum of the ML criterion ϵ (10). We prove this distribution to depend only on N and M leading to a θ -invariant regularization parameter as opposed to previous works that are θ -dependent [11].

V. DISTRIBUTION OF THE ML CRITERION MINIMUM

In this section, we derive the distribution of the minimum of the ML criterion under complex Gaussian non-circular noise. We prove the following theorem:

Theorem 1: Let us assume that K is sufficiently large so that:

- $\mathcal{H}1)$ The estimated whitening matrix is close to its asymptotic equivalent $\widehat{\mathbf{W}} \approx \mathbf{W}$.
- $\mathcal{H}2)$ The global minimum of J_{ML} is reached for $\hat{\Theta}$ which lies in the neighborhood of Θ such that $\hat{\gamma} \approx \gamma_0$. Therefore, ϵ can be approximated:

$$\epsilon = \mathbf{y}^H \mathbf{\Pi}^\perp(\hat{\Theta}) \mathbf{y} \approx \mathbf{y}^H \mathbf{\Pi}^\perp(\Theta) \mathbf{y} = \tilde{\epsilon} \quad (18)$$

where $\mathbf{\Pi}^\perp(\Theta)$ is the noise projector computed using both true directions and whitening matrix.

Then, the distribution of ϵ can be approximated by the distribution of $\tilde{\epsilon}$ which follows a χ^2 law with $N^2 - M$ degrees of freedom.

To prove theorem 1, let us first introduce the following properties of the permutation matrix \mathbf{K} :

Property 1: For any square matrices $\mathbf{M}_1, \mathbf{M}_2$ of size N^2 \mathbf{K} satisfies [17]:

$$\begin{aligned} \mathbf{K}^T &= \mathbf{K}, \mathbf{K}^2 = \mathbf{I}_{N^2}, \mathbf{K}^{-1} = \mathbf{K} \\ \mathbf{K}(\mathbf{M}_1 \otimes \mathbf{M}_2) &= (\mathbf{M}_1 \otimes \mathbf{M}_2) \mathbf{K}^{-1} = (\mathbf{M}_1 \otimes \mathbf{M}_2) \mathbf{K} \end{aligned} \quad (19)$$

Proof of theorem 1: To derive the distribution of ϵ , let us first transform \mathbf{y} into a real random vector by separating its real and imaginary parts. The corresponding real vector \mathbf{z} has size $2N^2$ and is given by:

$$\mathbf{z} = \begin{bmatrix} \Re\{\mathbf{y}\} \\ \Im\{\mathbf{y}\} \end{bmatrix} = \mathbf{H} \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \quad (20)$$

where

$$\mathbf{H} = \mathbf{Q} \otimes \mathbf{I}_{N^2} \quad \text{and} \quad \mathbf{Q} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -j & j \end{bmatrix} \quad (21)$$

Using the properties of the permutation matrix (19) and the Hermitian symmetry of $\widehat{\mathbf{R}}_x - \sigma^2 \mathbf{I}_{N^2}$, we have: $\mathbf{y}^* = \mathbf{K} \mathbf{y}$ leading to:

$$\mathbf{z} = \mathbf{H} \begin{bmatrix} \mathbf{I}_{N^2} \\ \mathbf{K} \end{bmatrix} \mathbf{y} = \mathbf{H} \mathbf{U} \mathbf{y} = \widetilde{\mathbf{B}}(\Theta) \gamma_0 + \mathbf{H} \mathbf{U} \delta_w \quad (22)$$

where $\widetilde{\mathbf{B}}(\Theta) = \mathbf{H} \mathbf{U} \mathbf{W} \mathbf{B}(\Theta)$. The noise in (22) is now a real random vector of law $\mathcal{N}(\mathbf{0}_{2N^2 \times 1}, \mathbf{\Sigma})$ with covariance matrix:

$$\mathbf{\Sigma} = \mathbb{E} \left[(\mathbf{H} \mathbf{U} \delta_w) (\mathbf{H} \mathbf{U} \delta_w)^H \right] = \mathbf{H} \mathbf{U} \mathbf{U}^H \mathbf{H}^H \quad (23)$$

since $\mathbb{E} \left[\delta_w \delta_w^H \right] = \mathbf{I}_{N^2}$ (7, 8). Using (19), we have:

$$\mathbf{H} \mathbf{U} = (\mathbf{Q} \otimes \mathbf{I}_{N^2}) \begin{bmatrix} \mathbf{I}_{N^2} \\ \mathbf{K} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_N^2 + \mathbf{K} \\ j(\mathbf{K} - \mathbf{I}_{N^2}) \end{bmatrix} \quad (24)$$

and inserting (24) in (22) leads to:

$$\Sigma = \frac{1}{2} \begin{bmatrix} \mathbf{I}_{N^2} + \mathbf{K} & \mathbf{0}_{N^2} \\ \mathbf{0}_{N^2} & \mathbf{I}_{N^2} - \mathbf{K} \end{bmatrix} \quad (25)$$

It can easily be verified that Σ is idempotent ($\Sigma^2 = \Sigma$) and has Hermitian symmetry. Hence, Σ is an orthogonal projection of rank N^2 . Given that its eigenvalues are 0 and 1 with multiplicity N^2 , the Eigenvalue Decomposition (EVD) of Σ gives $\Sigma = \mathbf{E}_s \mathbf{E}_s^T$ where \mathbf{E}_s contains the eigenvectors extracted from the N^2 non-null eigenvalues of Σ . Consequently, the noise $\mathbf{H}\mathbf{U}\delta_w$ in (22) can be rewritten using a white Gaussian noise vector $\tilde{\mathbf{n}} \sim \mathcal{N}(\mathbf{0}_{N^2 \times 1}, \mathbf{I}_{N^2})$:

$$\mathbf{z} = \tilde{\mathbf{B}}(\Theta)\gamma_0 + \mathbf{E}_s \tilde{\mathbf{n}} \quad (26)$$

Projecting \mathbf{z} on the columns of \mathbf{E}_s yields:

$$\mathbf{z}_s = \mathbf{E}_s^T \mathbf{z} = \mathbf{E}_s^T \tilde{\mathbf{B}}(\Theta)\gamma_0 + \tilde{\mathbf{n}} \quad (27)$$

since \mathbf{E}_s is a semi-orthogonal matrix.

Finally, $\tilde{\epsilon}$ is obtained as the the square norm of the projection of \mathbf{z}_s on the orthogonal subspace of $\tilde{\mathbf{B}}(\Theta)$. The corresponding projector is denoted $\Pi^\perp(\mathbf{E}_s \tilde{\mathbf{B}}(\Theta))$ and has rank $N^2 - M$ since $\mathbf{W}\mathbf{B}(\Theta)$ has rank M and $\mathbf{H}, \mathbf{U}, \mathbf{E}_s$ are full rank matrices. EVD yields $\Pi^\perp(\mathbf{E}_s \tilde{\mathbf{B}}(\Theta)) = \mathbf{I}_{N^2} - (\mathbf{E}_s \tilde{\mathbf{B}}(\Theta))^\# (\mathbf{E}_s \tilde{\mathbf{B}}(\Theta)) = \mathbf{F}_s \mathbf{F}_s^T$ with \mathbf{F}_s the semi-orthogonal matrix containing eigenvectors associated to $N^2 - M$ non-null eigenvalues. Using that the non-null eigenvalues are all equal to one, we have:

$$\tilde{\epsilon} = \mathbf{z}_s^T \Pi^\perp(\mathbf{E}_s \tilde{\mathbf{B}}(\Theta)) \mathbf{z}_s = (\mathbf{F}_s \tilde{\mathbf{n}})^T (\mathbf{F}_s \tilde{\mathbf{n}}) = \sum_{m=1}^{N^2-M} \tilde{n}_m^2 \quad (28)$$

Given that $\tilde{\epsilon}$ is a sum of $N^2 - M$ i.i.d Gaussian random variables, we conclude that $\tilde{\epsilon} \sim \chi_{N^2-M}^2$ which is a θ -invariant distribution. ■

VI. NUMERICAL EXPERIMENTS

We consider an array of $N = 6$ antennas with 5 antennas distributed around a circle of radius $0.8\lambda_0$ where λ_0 denotes the wavelength and one central sensor. $M = 2$ sources of directions $\theta_1 = 180^\circ, \theta_2 = 200^\circ$ and SNR = 0 dB impinge on this array. In simulations, ℓ_0 -norm in problem (13) is replaced by the Continuous Exact ℓ_0 (CELO) functional [18] to alleviate some of the optimization issues inherent to the ℓ_0 -norm. The corresponding criterion has less local minimums while having the same global minimum as J_{ℓ_0} . Hence, the previous value of λ remains valid for the CELO criterion. Finally, (13) is solved using the Forward-Backward Splitting (FBS) algorithm [19].

A. Distribution of ϵ

We first investigate the validity of theorem (1). The Cumulative Distribution Function (CDF) of ϵ is estimated for $K = 10^3$ (referred as an asymptotic case where $\tilde{\mathbf{W}} \approx \mathbf{W}$ and $\Pi^\perp(\tilde{\Theta}) \approx \Pi^\perp(\Theta)$), $K = 30, 100$ and $K = 200$. The ML estimate $\hat{\Theta}$ of the DOA is computed using the Gauss-Newton method [14] applied on (8). Results are represented on Fig.1.

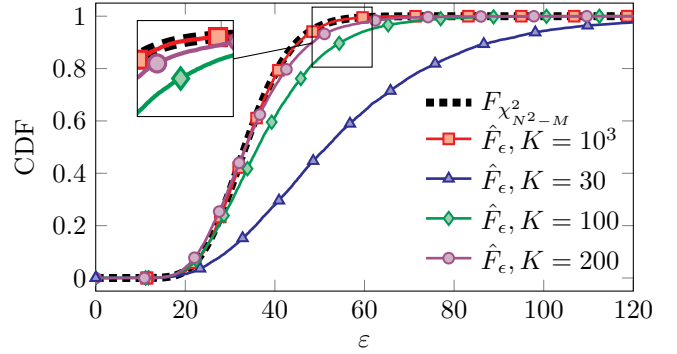


Fig. 1. Estimated CDF of ϵ for $K = 10^3$ (asymptotic case), $K = 30, 100$ and $K = 200$.

In asymptotic conditions ($K = 10^3$), the estimated CDF matches with the χ_2 CDF validating the results of theorem 1. Lowering K degrades the estimates of \mathbf{W} and Θ which shifts ϵ distribution towards larger values. For $K = 30$, the minimum of J_{ML} is reached for $\hat{\Theta} \neq \Theta$ which gives larger values of J_{ML} . Moreover, in this specific case \mathbf{y} is not Gaussian. Increasing K makes $\hat{\Theta}$ closer to Θ . Hence, ϵ decreases till it is approximately equal to $\tilde{\epsilon}$ leading to $\hat{F}_\epsilon \approx F_{\chi^2(N^2-M)}$ for $K \geq 200$. Thus, hypothesis $\mathcal{H}1$ and $\mathcal{H}2$ are valid leading to $\Pi^\perp(\hat{\Theta}) \approx \Pi^\perp(\Theta)$ for $K \geq 200$. In what follows, we set the number of samples to $K = 200$.

B. Influence of λ

As proposed in section IV, λ could be chosen using the CDF of ϵ such that condition (14) holds. We propose:

$$\lambda = \frac{1}{2} F_\epsilon^{-1}(\eta) \quad (29)$$

where $F_\epsilon^{-1}(\eta)$ is the inverse CDF of ϵ evaluated for probability η . To choose η , the sparsity of the estimated spectrum is evaluated for multiple λ values through a Monte-Carlo simulation. To compute sparsity, ℓ_0 -norm is approximated by a thresholding function with a threshold of 10^{-6} . The left plot of Fig.2 depicts the theoretical CDF of λ and the estimated CDF of λ_2^- and λ_2^+ on 10000 independent experiments. Using this figure, we choose η so that $\lambda_2^-(0.95) \leq \lambda \leq \lambda_2^+(0.05)$ where $\lambda_2^-(0.95)$ and $\lambda_2^+(0.05)$ are respectively upper and lower bounds on the distributions of (15). Additional simulations show that $\lambda_2^+(0.05)$ remains approximately constant while $\lambda_2^-(0.05)$ decreases for smaller values of $\Delta\theta = |\theta_1 - \theta_2|$. Hence, we propose to pick $\eta = 0.05$ which leads gives λ close $\lambda_2^-(0.95)$. This choice is then robust for several values of $\Delta\theta$.

On the the right plot of Fig.2, we represented the sparsity of the sparse spectrum estimate $\hat{\gamma}$ through the FBS algorithm as a function of λ . The proposed choice for λ belongs to the admissible interval I_2 [11] and leads to 2-sparse solutions.

C. Statistical performance

Finally, we verify that, for $\lambda = \frac{1}{2} F_{\chi^2(34)}^{-1}(0.05)$ ML and sparse methods lead to the same statistical performance. For

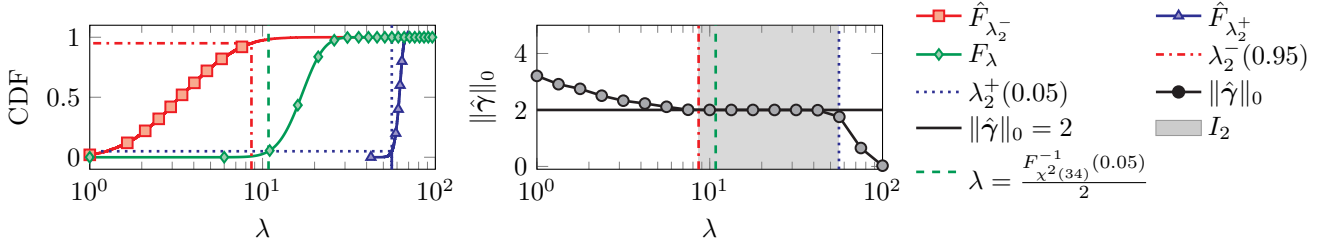


Fig. 2. Left: CDFs of λ_2^- , λ_2^+ and $\lambda = \frac{1}{2}\epsilon$. Right: Sparsity of $\hat{\gamma}$ as a function of λ .

that, we conducted 10000 independent Monte-Carlo experiments for SNR values between -20 dB and 20 dB. The sources are considered resolved if two peaks are detected and $\max\{|\hat{\theta}_1 - \theta_1|, |\hat{\theta}_2 - \theta_2|\} < 10^\circ$ where 10° corresponds to the half beamwidth of the virtual array. The corresponding number of correct detections is denoted Q . The Root Mean Square Error (RMSE) is computed for each direction on the correct detections as $\text{RMSE}(\theta) = \frac{1}{Q} \sqrt{\sum_{q=1}^Q (\hat{\theta}_{m,q} - \theta_m)^2}$. Results are represented on Fig.3. For $\text{SNR} \geq -4$ dB, ML and sparse estimators have the same performance which confirms that the chosen λ ensure the equivalence between both methods.

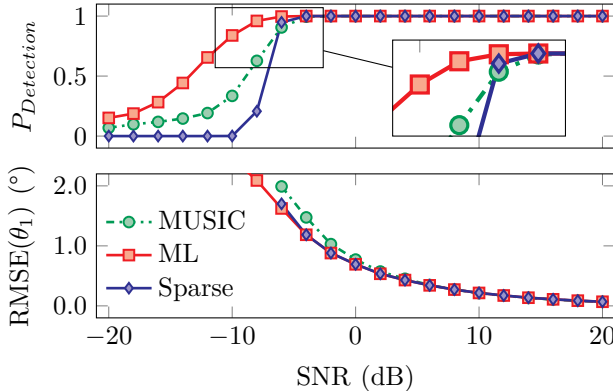


Fig. 3. Probability of detection (top) and RMSE (bottom). Given that the two sources have the same power, results are only represented for direction θ_1 .

VII. CONCLUSION

In this paper, we performed a theoretical analysis of equivalence between ML and sparse DOA estimators. We proposed to choose the regularization parameter λ using the distribution of the minimum of ML criterion. For that, the proposed λ is independent of the sources directions. It only requires an upper bound on the number of sources. Numerical simulations confirmed that under an oracle grid, the ML and sparse method have equal performance. For a practical implementation of the ML through sparse methods, off-grid techniques must be used. Moreover, our analysis should be extended to $M > 2$ with potentially correlated sources. This is an ongoing work.

REFERENCES

- [1] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *Signal Processing Magazine, IEEE*, vol. 13, pp. 67–94, 08 1996.
- [2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [3] G. Biennu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 5, pp. 1235–1248, 1983.
- [4] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," 2017.
- [5] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [6] M. Atashbar and M. H. Kahaei, "Direction-of-arrival estimation using amlss method," *IEEE Latin America Transactions*, vol. 10, no. 5, pp. 2053–2058, 2012.
- [7] Z. He, Q. Liu, L. Jin, and S. Ouyang, "Low complexity method for doa estimation using array covariance matrix sparse representation," *Electronics Letters*, vol. 49, no. 3, pp. 228–230, 2013.
- [8] W. Cui, T. Qian, and J. Tian, "Enhanced covariances matrix sparse representation method for doa estimation," *Electronics Letters*, vol. 51, no. 16, pp. 1288–1290.
- [9] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4489–4493, 2011.
- [10] N. R. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction)," *The Annals of mathematical statistics*, vol. 34, no. 1, pp. 152–177, 1963.
- [11] A. Delmer, A. Ferréol, and P. Larzabal, "On regularization parameter for 10-sparse covariance fitting based doa estimation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4552–4556.
- [12] P. Chevalier, L. Albera, A. Ferréol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1254–1271, 2005.
- [13] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez, "Asymptotic properties of robust complex covariance matrix estimates," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3348–3356, 2013.
- [14] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, *Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, pp. 99–151.
- [15] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 09 1994.
- [16] P. C. Hansen, *The L-Curve and Its Use in the Numerical Treatment of Inverse Problems*, 01 2001, vol. 4, pp. 119–142.
- [17] J. Magnus and H. Neudecker, "The commutation matrix: Some properties and applications," *Annals of Statistics*, vol. 7, no. 2, pp. 381–394, 1979, pagination: 14.
- [18] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A Continuous Exact 10 penalty (CEL0) for least squares regularized problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. pp. 1607–1639 (33 pages), Jul. 2015.
- [19] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," 2010.