ML-DoA Estimation using a Sparse Representation of Array Covariance with a non-standard noise

Thomas Aussaguès, Anne Ferréol, Alice Delmer, Pascal Larzabal

ML-DoA Estimation using a Sparse Representation of Array Covariance with a non-standard noise

Thomas Aussaguès[a,b,*], Anne Ferréol[a,b], Alice Delmer[a], Pascal Larzabal[b,]

[a] *Thales, 4 avenue des Louvresses, Gennevilliers, 92230, Hauts-de-Seine, France*
[b] *SATIE, Université Paris-Saclay, UMR CNRS 8029, 4 avenue des Sciences, Gif-sur-Yvette, 91190, Essonne, France*

Abstract

Maximum Likelihood (ML) Direction-of-Arrival (DoA) on the vectorized covariance matrix model (VCMM), relying on a Virtual Array (VA) of antennas, exhibits enhanced ability to separate closely spaced sources. Due to the finite number of snapshots, the VCMM observation is corrupted by a non-white and non-circular Gaussian noise resulting in an intricate ML criterion. To address this issue, this paper introduces a novel two-stage transform that turns the initial non-white and non-circular Gaussian noise into a real and white Gaussian noise. Following this, the ML estimator is formulated for the transformed model and the corresponding Cramér-Rao Lower Bound (CRLB) is derived.

Unfortunately, the ML implementation involves intractable multi-dimensional and highly non-linear non-convex optimization.

This paper introduces a novel sparse DoA estimator that implements the ML using the proposed two-stage transform. This transform is shown to significantly simplify the sparse estimator implementation. To quantify the transform effects, the problem conditioning is derived and shown to be consequently improved after the transform. Numerical simulations showcase performance improvements of the sparse DoA estimator after the two-stage transform in severe scenarios with 2 and 3 closely spaced sources using a 4-elements array.

*Corresponding author
Email address: thomas.aussagues@ens-paris-saclay.fr (Thomas Aussaguès)

1. Introduction

Direction-of-Arrival (DoA) estimation is typical signal processing problem arising in numerous applications such as radar, sonar or telecommunications. Throughout last decades, a myriad of estimation techniques has been proposed [1, 2]. Among these, Capon's beamformer [3], subspace identification based methods like MUSIC [4, 5] or ESPRIT [6] and the Maximum Likelihood (ML) estimator [7, 8, 9] emerge as the most popular estimators. Despite their popularity, above-mentioned methods suffer from serious limitations. For instance, MUSIC fails in severe scenarios with highly correlated sources or few time snapshots. The ML estimator, while capable of handling correlated sources and being statistically efficient at high Signal-to-Noise Ratio (SNR) under white Gaussian noise [10], is seldom used in practice. This is primarily because ML estimation requires solving an intractable, multi-dimensional, highly non-convex optimization problem with numerous local minima [9]. Moreover, all aforementioned methods share a common limitation: the maximal number of identifiable sources is limited to $N-1$ using an N-element array.

To tackle some of the previously mentioned issues, one can leverage a VA, with space diversity only, of at most N^2-N non-redundant antennas. This approach offers increased number of identifiable sources and enhanced ability to separate closely spaced sources [11]. The VA can be accessed through extensions of MUSIC to higher order statistics [12, 13] assuming non-Gaussian emitters. Alternatively, the VCMM can be employed to access the VA [14]. This novel model only necessitates second order statistics, which exhibits better convergence than higher order statistics, of the observation without any assumptions on the sources signals statistics.

Recently, the signal processing community extensively investigated the topic of sparse signal representation to address the shortcomings classical methods limitations [15, 2]. Several authors experimentally demonstrated the superiority of sparse-based DoA estimators in challenging scenarios [16, 14, 17, 18] compared to traditional techniques such as MUSIC. Specifically, these estimators [14, 17,

18] rely on a sparse representation of the vectorized covariance matrix thus levering the VA and thereby enhancing performance.

Using this sparse representation, the DoAs are retrieved through the minimization of a ℓ_0-regularized sparse criterion parametrized by λ the regularization parameter [18]. Despite the VA advantages, the processing of poorly separated sources remains challenging for a sparse estimator due the spatial correlation between directions, which complicates the criterion optimization [19]. In [20], a novel θ-invariant regularization parameter choice is introduced to ensure equivalence between sparse and ML DoA estimators under white Gaussian noise. As a side effect, this approach enables the theoretical characterization of the sparse estimator by the way of ML performance. This equivalence relies on a two-stage transform that handles both the noise correlation and non-circularity. As outlined in [20], the VCMM is corrupted by a coloured noise vector due to the finite number of samples. The presence of coloured noise can significantly deteriorate performance [21, 22]. To address this, several authors have proposed modified estimators [23, 24, 25]. When sufficient data is available, a pre-whitening noise transform is applied transforming the initially coloured noise into white noise [9]. Such procedure is performed on the VCMM in [20] resulting in a real white Gaussian noise thus enabling equivalence with the ML under white Gaussian noise. Additionally, this transform modifies the dictionary matrix involved in the sparse criterion by decorrelating vectors associated to sources directions, thereby improving the resolvability of closely spaced sources [19].

This paper seeks to extend the findings of [19] and is three-fold. First, it introduces a two-stage transform of the VCMM that converts the initially complex non-circular and non-white Gaussian noise into a novel real standard white Gaussian noise. This transformation not only addresses noise correlation and non-circularity but also yields a simple deterministic ML criterion compared to the multi-term intricate criterion obtained under coloured noise. Using the transformed model, the Cramér-Rao Lower Bound is derived to characterize the asymptotic estimator performance.

In addition, the aforementioned two-stage transform also affects the sparse

criterion. The impacts of this transform on sparse criterion are discussed in [19]. Specifically, the problem conditioning is shown to be significantly improved after the transform, leading to substantial enhancements in the sparse criterion optimization. While [19] analyzed the spatial correlation coefficient, its formal relationship to the problem conditioning was reserved for future investigation. The present work establishes a formal connection between spatial correlation and problem conditioning.

Finally, the two-stage transform effects on the Continuous Exact ℓ_0 (CEL0) loss surface [26] are examined. As demonstrated in [27], while this penalty simplifies the criterion optimization by reducing the number of local minima induced by the ℓ_0 penalization, the CEL0 criterion exhibits flat minimums in the case of closely separated sources, thereby complicating their estimation. Through numerical simulations, the transform is shown to reduce the sparse criterion corridors hence enhancing the estimation of closely spaced sources.

This paper is organized as follows: section 2 introduces the VCMM and the proposed two-stage transform, resulting in a transformed model corrupted by a standard white Gaussian noise. Then, the ML estimator and the CRLB derivation are presented in section 3 for the transformed model under white Gaussian noise. Section 4 is devoted to the sparse estimator and its equivalence with the ML. In section 5, a theoretical investigation on the two-stage transform effects on the sparse criterion is conducted. The transformed dictionary is proven to have reduced correlation between sources directions thus enhancing the criterion optimization. Numerical illustrations of the criterion are provided to support our claims. Finally, in section 6, numerical simulations for scenarios with 2 and 3 sources are conducted to assess the performance enhancement of the proposed sparse DoA estimator.

Notations. *Upper-case and lower-case boldface Latin letters respectively denote matrices and vectors.* $(\cdot)^\star$ *denotes the complex conjugate,* $(\cdot)^T$ *the transpose and* $(\cdot)^H$ *the conjugate transpose of a vector or matrix.* $\otimes$ *and* $(\cdot)^\#$ *refer to the Kronecker product and the Moore-Penrose pseudo-inverse of a matrix. Lastly,*

$\mathbb{E}[\cdot]$ *denotes the temporal mean operator,* $\widehat{(\cdot)}$ *the estimate of a parameter and* $\mathbf{I}_N$ *the identity matrix of size* N*.*

2. Signal modelling, hypothesis and problem formulation

2.1. The vectorized covariance matrix model

Let M independent narrowband plane waves of directions $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_M\}$ impinge on an array of N antennas. The array output is then:

$$\mathbf{x}(t) = \sum_{m=1}^{M} \mathbf{a}(\theta_m)\mathbf{s}_m(t) + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_{\boldsymbol{\theta}}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_M))]$ is the steering matrix formed by the steering vectors $\mathbf{a}(\theta_m), 1 \leq m \leq M$, $\mathbf{s}_{\boldsymbol{\theta}}(t) = [s_1(t), \ldots, s_M(t)]^T$ the complex envelopes of the emitted signals which are considered unknown but deterministic and $\mathbf{n}(t)$ a complex circular Gaussian noise vector [28], independent of $\mathbf{s}_{\boldsymbol{\theta}}(t)$, with covariance matrix $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}^H(t)\right] = \sigma^2\mathbf{I}_N$.

The covariance matrix of (1) is thus:

$$\mathbf{R}_x = \mathbb{E}\left[\mathbf{x}(t)\mathbf{x}^H(t)\right] = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_N \tag{2}$$

where $\mathbf{R}_s = \mathbb{E}\left[\mathbf{s}_{\boldsymbol{\theta}}(t)\mathbf{s}_{\boldsymbol{\theta}}^H(t)\right]$ refers to the sources covariance matrix. Throughout this paper, sources are assumed to be temporally uncorrelated leading to a diagonal sources covariance matrix satisfying $\mathbf{R}_s = \mathrm{diag}(\boldsymbol{\gamma}_{\boldsymbol{\theta}})$ with $\boldsymbol{\gamma}_{\boldsymbol{\theta}} = [\gamma_1, \ldots, \gamma_M]^T$ the sources powers vectors and $\mathrm{diag}(\boldsymbol{x})$ the diagonal matrix whose diagonal is $\boldsymbol{x}$.

The use of (1) and (2) restricts the performance in severe scenarios (few array snapshots, low SNR or closely spaced sources). To improve theses limitations, the VCMM is employed. This model relies on an observation $\mathbf{r}$ of size N^2 associated to the output of a VA with N^2 sensors among which at most $N^2 - N$ are non-redundant. The VA has fewer sidelobes and a thinner mainlobe leading to an enhanced ability to resolve closely spaced sources [11].

Under the assumption of temporally uncorrelated emitters, the VCMM observation is [14]:

$$\mathbf{r} = \text{vec}\left(\mathbf{R}_x - \sigma^2\mathbf{I}_N\right) = \sum_{m=1}^{M} \mathbf{b}(\theta_m)\gamma_m = \mathbf{B}(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} \tag{3}$$

with $\text{vec}(\cdot)$ the column-wise vectorisation operator, $\mathbf{B}(\boldsymbol{\theta})$ the VA steering matrix formed by the vectors $\mathbf{b}(\theta_m) = \mathbf{a}^\star(\theta_m) \otimes \mathbf{a}(\theta_m)$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ defined above.

In practice, the true covariance matrix $\mathbf{R}_x$ is not accessible and is thus replaced by its corresponding ML estimate $\widehat{\mathbf{R}}_x$ obtained using K identically and independently distributed array snapshots $\mathbf{x}(t_k), 1 \leq k \leq K$. Assuming temporally white noise on $\mathbf{x}(t)$ ($\mathbb{E}\left[\mathbf{n}^H(t_i)\mathbf{n}(t_j)\right] = 0$ for $i \neq j$), $\widehat{\mathbf{R}}_x$ can be decomposed as:

$$\widehat{\mathbf{R}}_x = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}(t_k)\mathbf{x}^H(t_k) = \mathbf{R}_x + \boldsymbol{\Delta}\mathbf{R}_x \tag{4}$$

where $\boldsymbol{\Delta}\mathbf{R}_x$ is a random matrix following a complex Wishart distribution [29]. Consequently, model (3) is corrupted by the covariance matrix estimation error $\boldsymbol{\delta} = \text{vec}\left(\Delta\mathbf{R}_x\right)$ leading to:

$$\mathbf{r} = \text{vec}\left(\widehat{\mathbf{R}}_x - \sigma^2\mathbf{I}_N\right) = \mathbf{B}(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \boldsymbol{\delta} \tag{5}$$

where $\boldsymbol{\delta}$ refers to the noise on the observation $\mathbf{r}$.

By the way of the central limit theorem and [30], the complex Wishart distribution of the covariance matrix estimation error $\Delta\mathbf{R}_x$ is asymptotically (with respect to K) a complex Gaussian distribution. The vectorization of (4) then yields a corresponding complex Gaussian noise vector (5) of law $\mathbb{C}\mathcal{N}\left(\mathbf{0}_{N^2\times 1}, \boldsymbol{\Gamma}, \mathbf{C}\right)$ [28] with the following moments [31]:

$$\boldsymbol{\Gamma} = \mathbb{E}\left[\boldsymbol{\delta}\boldsymbol{\delta}^H\right] = \frac{1}{K}\left(\mathbf{R}_x^T \otimes \mathbf{R}_x\right) \quad \mathbf{C} = \mathbb{E}\left[\boldsymbol{\delta}\boldsymbol{\delta}^T\right] = \boldsymbol{\Gamma}\mathbf{K} \tag{6}$$

where $\boldsymbol{\Gamma}$ and $\mathbf{C}$ respectively denote the covariance matrix and the pseudo-covariance matrix of the covariance matrix estimation vectorized error $\boldsymbol{\delta}$ and $\mathbf{K}$ is the permutation matrix such that $\text{vec}(\boldsymbol{\mathcal{M}}^T) = \mathbf{K}\text{vec}(\boldsymbol{\mathcal{M}})$ for any square matrix $\boldsymbol{\mathcal{M}} \in \mathbb{C}^{N^2\times N^2}$ [32].

2.2. Two-stage transform of the observation $\mathbf{r}$

As outlined in section 2, the VCMM observation (5) is corrupted by a complex non-white and non-circular Gaussian noise vector $\boldsymbol{\delta}$ (6). This yields an intricate multi-term log-likelihood function which could be simplified in the case of white Gaussian noise.

To this end, this section introduces a novel two-stage transform exploiting only the noise second order statistics. The transform purpose is to turn the initially non-white and non-circular complex Gaussian noise into a real white Gaussian noise.

First, the noise is whitened as described in 2.2.1 through the whitening matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\mathbf{r}$ where $\mathbf{y}$ it the resulting observation. After the whitening, the noise on $\mathbf{y}$ is white but remains non-circular.

Then, the non-circularity is handled in 2.2.2 through a decomposition of the real and imaginary parts of the observation. The corresponding observation noise is non-white since its covariance matrix differs from the identity. Thus, a final whitening step is performed. These steps are synthesised by the matrix $\mathbf{T}$

The resulting real observation after the two-stage transform is denoted $\mathbf{z} = \mathbf{T}\mathbf{y} = \mathbf{T}\mathbf{W}\mathbf{r}$ and is corrupted by a real white Gaussian noise.

2.2.1. Noise whitening

From eq. (6), it follows that $\boldsymbol{\delta}$ is non-white. To transform its covariance matrix $\boldsymbol{\Gamma}$ into $\mathbf{I}_{N^2}$, eq. (5) is pre-multiplied by $\mathbf{W}$ the $N^2 \times N^2$ whitening matrix as done in [20, 19]:

$$\mathbf{y} = \mathbf{W}\mathbf{r} = \mathbf{W}\mathbf{B}(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \mathbf{W}\boldsymbol{\delta} = \mathbf{B}_w(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \boldsymbol{\delta}_w \tag{7}$$

where $\mathbf{W} = \boldsymbol{\Gamma}^{-1/2} = \sqrt{K}\left(\mathbf{R}_x^{-T/2} \otimes \mathbf{R}_x^{-1/2}\right)$, $\mathbf{B}_w(\boldsymbol{\theta}) = \mathbf{W}\mathbf{B}(\boldsymbol{\theta})$ denotes the dictionary obtained after the whitening and $\boldsymbol{\delta}_w = \mathbf{W}\boldsymbol{\delta}$ the whitened noise.

After whitening, the noise $\boldsymbol{\delta}_w$ is thus white but yet non-circular, such that:

$$\boldsymbol{\Gamma}_w = \mathbb{E}\left[\boldsymbol{\delta}_w\boldsymbol{\delta}_w^H\right] = \mathbf{I}_{N^2} \quad \mathbf{C}_w = \mathbb{E}\left[\boldsymbol{\delta}_w\boldsymbol{\delta}_w^T\right] = \mathbf{W}\boldsymbol{\Gamma}\mathbf{K}\mathbf{W}^T \tag{8}$$

where the matrix $\mathbf{C}_w$ can be further simplified using both permutation matrix [32] and Kronecker product [33] properties:

$$\begin{aligned}\mathbf{C}_w &= \left(\mathbf{R}_x^{-T/2} \otimes \mathbf{R}_x^{-1/2}\right)\left(\mathbf{R}_x^{T} \otimes \mathbf{R}_x^{T}\right)\mathbf{K}\left(\mathbf{R}_x^{-T/2} \otimes \mathbf{R}_x^{-1/2}\right)^T \\ &= \left(\mathbf{R}_x^{T/2} \otimes \mathbf{R}_x^{1/2}\right)\mathbf{K}\left(\mathbf{R}_x^{-1/2} \otimes \mathbf{R}_x^{-T/2}\right) \\ &= \left(\mathbf{R}_x^{T/2} \otimes \mathbf{R}_x^{1/2}\right)\left(\mathbf{R}_x^{-T/2} \otimes \mathbf{R}_x^{-1/2}\right)\mathbf{K} \\ &= \mathbf{K}\end{aligned} \tag{9}$$

since:

$$\mathbf{K}\left(\boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}}\right) = \left(\boldsymbol{\mathcal{B}} \otimes \boldsymbol{\mathcal{A}}\right)\mathbf{K} \quad \left(\boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}}\right)^T = \boldsymbol{\mathcal{A}}^T \otimes \boldsymbol{\mathcal{B}}^T \quad \left(\boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}}\right)^{-1} = \boldsymbol{\mathcal{A}}^{-1} \otimes \boldsymbol{\mathcal{B}}^{-1} \tag{10}$$

for any conformable matrices $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$.

Remark 1. *In practice, the whitening matrix is replaced by its corresponding estimate* $\widehat{\mathbf{W}} = \sqrt{K}\left(\widehat{\mathbf{R}}_x^{-T/2} \otimes \widehat{\mathbf{R}}_x^{-1/2}\right)$.

Remark 2. *Using the vectorization property of the Kronecker product (10), the whitened observation (7) can be rewritten as:*

$$\mathbf{y} = \sqrt{K}\mathrm{vec}\left(\mathbf{I}_N - \sigma^2\widehat{\mathbf{R}}_x^{-1}\right) \tag{11}$$

showing that the pre-whitening noise transform preserves (7) the DoA information as it is contained within $\widehat{\mathbf{R}}_x^{-1}$.

2.2.2. Transformation of the complex non-circular noise into a real white noise

The purpose of this paragraph is to take into account the non-circularity and to transform the non-circular noise $\boldsymbol{\delta}$ into a real white noise $\tilde{\mathbf{n}}$. The following transforms yield a novel observation $\mathbf{z} = \mathbf{T}\mathbf{y}$. Indeed, eq. (8) and (9) show that although the noise $\boldsymbol{\delta}_w$ is white after the whitening transform (7), it remains non-circular as $\mathbf{C}_w \neq \mathbf{0}_{N^2}$.

Non-circularity. The conventional approach to handle the noise non-circularity is to concatenate both real and imaginary parts of $\mathbf{y}$ (respectively $\Re\{\mathbf{y}\}$ and

$\Im\{\mathbf{y}\}$) to form a novel real augmented observation $\tilde{\mathbf{y}}$ of size $2N^2 \times 1$:

$$\tilde{\mathbf{y}} = \begin{bmatrix} \Re\{\mathbf{y}\} \\ \Im\{\mathbf{y}\} \end{bmatrix} = \mathbf{H} \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} \tag{12}$$

where:

$$\mathbf{H} = \mathbf{Q} \otimes \mathbf{I}_{N^2} \quad \text{and} \quad \mathbf{Q} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -j & j \end{bmatrix} \tag{13}$$

and $\mathbf{y} = \mathbf{W}\mathbf{r}$. Using both properties of the permutation matrix (10) and the Hermitian symmetry of $\widehat{\mathbf{R}}_x - \sigma^2\mathbf{I}_N$, we have $\mathbf{y}^\star = \mathbf{K}\mathbf{y}$ leading to:

$$\tilde{\mathbf{y}} = \mathbf{H} \begin{bmatrix} \mathbf{I}_{N^2} \\ \mathbf{K} \end{bmatrix} \mathbf{y} = \mathbf{H}\mathbf{U}\mathbf{y} = \mathbf{H}\mathbf{U}\mathbf{B}_w(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \mathbf{H}\mathbf{U}\boldsymbol{\delta}_w \tag{14}$$

The noise in (14) is now a real random vector of law $\mathcal{N}(\mathbf{0}_{2N^2\times 1}, \boldsymbol{\Sigma})$ with covariance matrix:

$$\boldsymbol{\Sigma} = \mathbb{E}\left[(\mathbf{H}\mathbf{U}\boldsymbol{\delta}_w) (\mathbf{H}\mathbf{U}\boldsymbol{\delta}_w)^H \right] = \mathbf{H}\mathbf{U}\mathbf{U}^H\mathbf{H}^H \tag{15}$$

since $\mathbb{E}\left[\boldsymbol{\delta}_w\boldsymbol{\delta}_w^H\right] = \mathbf{I}_{N^2}$ (8). Using $\mathbf{K}^2 = \mathbf{I}$ [32], we have:

$$\mathbf{H}\mathbf{U} = (\mathbf{Q} \otimes \mathbf{I}_{N^2}) \begin{bmatrix} \mathbf{I}_{N^2} \\ \mathbf{K} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_N^2 + \mathbf{K} \\ j(\mathbf{K} - \mathbf{I}_{N^2}) \end{bmatrix} \tag{16}$$

and inserting (16) in (15) leads to:

$$\boldsymbol{\Sigma} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_{N^2} + \mathbf{K} & \mathbf{0}_{N^2} \\ \mathbf{0}_{N^2} & \mathbf{I}_{N^2} - \mathbf{K} \end{bmatrix} \tag{17}$$

Final whitening step. After handling the non-circularity, the obtained real noise is non-white as $\boldsymbol{\Sigma} \neq \mathbf{I}_{N^2}$ (17) leading us to apply one final whitening step to obtain the observation $\mathbf{z}$ corrupted by a real white Gaussian noise.

The $2N^2 \times 2N^2$ covariance matrix $\boldsymbol{\Sigma}$ (17) is an orthogonal projection [34] since it is symmetric ($\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma}$) and idempotent ($\boldsymbol{\Sigma}^2 = \boldsymbol{\Sigma}$). $\boldsymbol{\Sigma}$ can rewritten through an eigenvalue decomposition $\boldsymbol{\Sigma} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^T$ where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues and $\mathbf{E}$ the matrix formed by the corresponding eigenvectors. Since $\boldsymbol{\Sigma}$ is an orthogonal projection, its first N^2 eigenvalues are equal

to 1 and the last N^2 to 0. Thus, the EVD can be simplified as $\mathbf{\Sigma} = \mathbf{E}_s \boldsymbol{E}_s^T$ with $\mathbf{E}_s$ the restriction of $\mathbf{E}$ to the eigenvectors associated to non-null eigenvalues all equal to one.

$\mathbf{E}_s$ is clearly a square root of $\mathbf{\Sigma}$. Thus, the noise $\mathbf{HU}\boldsymbol{\delta}_w$ of (14) is whitened by multiplying it with $\mathbf{E}_s^T$. This yields a novel observation $\mathbf{z}$ of size $N^2 \times 1$ (since $\mathbf{E}_s^T$ has size $2N^2 \times N^2$) obtained from the whitened observation $\mathbf{y}$ (7):

$$\mathbf{z} = \mathbf{E}_s^T \tilde{\mathbf{y}} = \mathbf{E}_s^T \mathbf{HUy} = \mathbf{E}_s^T \mathbf{HUB}_w(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \mathbf{E}_s^T \mathbf{HU}\boldsymbol{\delta}_w = \mathbf{B}_t(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \tilde{\mathbf{n}} \quad (18)$$

where the matrix $\mathbf{B}_t(\boldsymbol{\theta}) = \mathbf{TB}_w(\boldsymbol{\theta})$ is the transformed dictionary obtained using the transformation matrix $\mathbf{T} = \mathbf{E}_s\mathbf{HU}$ which summarizes all previous steps (decomposition of the real and imaginary parts eq. (14) and second whitening eq. (18)).

The transformed observation $\mathbf{z}$ (18) is now corrupted by the transformed noise, denoted $\tilde{\mathbf{n}} = \mathbf{T}\boldsymbol{\delta}_{\boldsymbol{w}}$, which is a real white Gaussian noise as desired since :

$$\mathbb{E}\left[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T\right] = \mathbf{E}_s^T \mathbf{HU}\left(\mathbf{E}_s^T \mathbf{HU}\right)^T = \mathbf{E}_s^T \mathbf{\Sigma} \mathbf{E}_s = \mathbf{E}_s^T \mathbf{E}_s \mathbf{E}_s^T \mathbf{E}_s \mathbf{I}_N^2 \quad (19)$$

as $\mathbf{E}_s^T \mathbf{E}_s = \mathbf{I}_{N^2}$. This concludes the two-stage transformation.

The transformed VCMM (18) is now contaminated by a real white Gaussian noise for which the ML estimator can now be easily formulated.

3. ML-DoA estimation on the transformed vectorized covariance matrix model z

3.1. Conditional Maximum Likelihood for white Gaussian noise

For the white Gaussian noise model (18), the conditional ML [7, 8, 9] estimates of the true DoAs $\boldsymbol{\theta}$ are obtained using the following estimator:

$$\begin{aligned} \mathcal{J}_{ML}(\boldsymbol{\varphi}) &= \mathrm{tr}\left(\mathbf{\Pi}^{\perp}(\boldsymbol{\varphi})\mathbf{z}\mathbf{z}^T\right) = \mathbf{z}^T \mathbf{\Pi}^{\perp}(\boldsymbol{\varphi})\mathbf{z} \\ \widehat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\varphi} \in \mathbb{R}^M} \mathcal{J}_{ML}(\boldsymbol{\varphi}) \end{aligned} \quad (20)$$

where $\mathbf{\Pi}^{\perp}(\boldsymbol{\varphi}) = \mathbf{I}_{N^2} - \mathbf{B}_t(\boldsymbol{\varphi})\mathbf{B}_t^{\#}(\boldsymbol{\varphi})$ is the noise projector computed for candidate directions $\boldsymbol{\varphi}$ and $\mathrm{tr}(\cdot)$ the trace operator.

Although its statistical efficiency as it achieves the Cramér-Rao Lower Bound at high SNR [10], the ML requires computationally intractable optimization since $\mathcal{J}_{ML}$ is an M-dimensional non-linear and non-convex criterion with numerous local minima [9]. As the number of local minima increases with the number of sources, the ML is confined to small number of sources.

3.2. Derivation of the Conditional Cramér-Rao Lower Bound based on the transformed model $\mathbf{z}$

The CRLB is a lower bound on the variance of all unbiased estimators of a parameters vector $\boldsymbol{\alpha}$ such that:

$$\text{CRLB}(\alpha_i) \leq \text{Var}(\alpha_i) = \left[\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\alpha})\right]_{ii} = \left[\mathbb{E}\left[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T\right]\right]_{ii} \tag{21}$$

with $\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha})$ the Fisher information matrix. For the special case of the multivariate real Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\alpha}), \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$, there exists a closed-form formula form the (i,j) entry of Fisher information matrix referred as the Slepian-Bangs formula [35, 36]:

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha})]_{ij} = \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\alpha})}{\partial \alpha_i} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\alpha})}{\partial \alpha_j} + \frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}) \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\alpha})}{\partial \alpha_i} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}) \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\alpha})}{\partial \alpha_j}\right) \tag{22}$$

Applying (22) to the deterministic model (18) with:

$$\begin{aligned} \boldsymbol{\alpha}^T &= \begin{bmatrix} \boldsymbol{\theta}^T & \boldsymbol{\gamma}_{\boldsymbol{\theta}}^T \end{bmatrix}^T = \begin{bmatrix} \theta_1 & \dots & \theta_M & \gamma_1 & \dots & \gamma_M \end{bmatrix}^T \\ \boldsymbol{\mu}(\boldsymbol{\alpha}) &= \mathbf{B}_t(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} \\ \boldsymbol{\Sigma}(\boldsymbol{\alpha}) &= \mathbf{I}_N \end{aligned} \tag{23}$$

yields:

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha})]_{ij} = \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\alpha})}{\partial \alpha_j} \quad \text{and} \quad \boldsymbol{\mathcal{I}}(\boldsymbol{\alpha}) = \boldsymbol{\nabla}\boldsymbol{\mu}(\boldsymbol{\alpha})\boldsymbol{\nabla}^T\boldsymbol{\mu}(\boldsymbol{\alpha}) \tag{24}$$

where straightforward derivations leads to $\boldsymbol{\nabla}\boldsymbol{\mu}(\boldsymbol{\alpha}) = \begin{bmatrix} \frac{\partial \mathbf{B}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \odot \boldsymbol{\gamma}_{\boldsymbol{\theta}}^T & \mathbf{B}_t(\boldsymbol{\theta}) \end{bmatrix}^T$ with $\odot$ the Hadamard product. Then, in the numerical simulations of section 6, the CRLB is numerically evaluated using eq. (21) and (24).

4. Sparse ML-DoA estimation on the vectorized covariance matrix model

To overcome the ML implementation pitfalls detailed in section 1, a sparse estimator based on (18) is employed. Equivalence under white Gaussian noise between this sparse estimator and the ML (20) has recently been established for proper hyperparameter selection [20].

4.1. Sparse modelling

Let us discretize the angular space using a grid of G pre-defined directions $\boldsymbol{\varphi} = \{\varphi_1, \ldots, \varphi_G\}$. Under the assumption that the sources directions $\boldsymbol{\theta}$ lie within $\boldsymbol{\varphi}$, eq. (18) can be rewritten using the sparse framework:

$$\mathbf{z} = \mathbf{B}_t(\boldsymbol{\varphi})\boldsymbol{\gamma}_0 + \tilde{\mathbf{n}} \tag{25}$$

where $\mathbf{B}_t(\boldsymbol{\varphi}) = \mathbf{T}\mathbf{B}_w(\boldsymbol{\varphi})$ with $\mathbf{B}_w(\boldsymbol{\varphi}) = \mathbf{W}[\mathbf{b}(\varphi_1), \ldots, \mathbf{b}(\varphi_G)]$ is an overcomplete dictionary of size $N^2 \times G, G \gg N^2$ and $\boldsymbol{\gamma}_0$ is an M-sparse vector having M non-zero components at directions satisfying $\varphi_g = \theta_m$. Finally, $\tilde{\mathbf{n}}$ is a real white Gaussian noise vector satisfying $\mathbb{E}\left[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T\right] = \mathbf{I}_{N^2}$.

4.2. Sparse estimation

The DoAs can be estimated from the sparse model (25) using the grid directions φ_g which are associated to non-null entries of the sparse vector $\boldsymbol{\gamma}_0$. Consequently, an estimate of $\boldsymbol{\gamma}_0$ is needed to estimate $\boldsymbol{\theta}$. Given that the G unknown coefficients of $\boldsymbol{\gamma}_0$ are estimated from an observation of length N^2, the problem is ill-posed and thus can not be resolved through the least squares minimization. The estimation (25) is addressed by leveraging the sparsity prior to ensure the uniqueness of the solution. The sparsity is thus enforced through the addition of a ℓ_0-regularizer leading to the following ℓ_0-regularized objective [18, 20]:

$$\min_{\boldsymbol{\gamma} \in \mathbb{C}^G} \left\{ \mathcal{J}_{\ell_0}(\lambda, \boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{z} - \mathbf{B}_t(\boldsymbol{\varphi})\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_0 \right\} \tag{26}$$

where $\lambda > 0$ refers to the regularization parameter which balances the solution sparsity towards data fidelity.

Remark 3. *The dictionary vectors* $\mathbf{b}_t(\varphi_g)$ *contained in* $\mathbf{B}_t(\boldsymbol{\varphi})$ *are normalized such that* $\|\mathbf{b}_t(\varphi_g)\|_2 = 1$ *to ensure that all directions have equal energy.*

4.3. Selection of the regularization parameter

In [20], authors applied the work of [37, 18] to the transformed model (18). After exploiting the statistics of the minimum of the ML criterion (20) for K sufficiently large, they proposed to pick λ as:

$$\lambda = \frac{1}{2}\mathcal{F}_{\chi_2(N^2-M)}(\eta) \tag{27}$$

where $\mathcal{F}_{\chi_2(N^2-M)}$ is the $\chi_2(N^2 - M)$ distribution with $N^2 - M$ degrees of freedom cumulative distribution function and η a probability (typically set to 5×10^{-2} [20]). Furthermore, it was shown that this choice ensures equivalence between sparse and ML estimator after the transformation of subsection 2.2 *ie.* both criteria have the same global minimizer.

5. Whitening effects for sparse ML-DOA estimation on the transformed observation z

In this section, the consequences of the two-stage transform on the dictionary matrix are investigated with $M = 2$ impinging sources.

5.1. Enhancement of the problem conditioning through whitening

The transform affects the sparse criterion (18) by transforming the dictionary which controls the criterion shape. To measure the transform effects on the dictionary vectors, let us introduce the spatial correlation function:

$$r_{\mathbf{E}}(\varphi_i, \varphi_j) = \frac{|\mathbf{e}^H(\varphi_i)\mathbf{e}(\varphi_j)|}{\|\mathbf{e}(\varphi_i)\|_2\|\mathbf{e}(\varphi_j)\|_2} \tag{28}$$

which measures the spatial correlation between directions φ_i and φ_j of a given dictionary $\mathbf{E} = [\mathbf{e}(\varphi_1), \ldots, \mathbf{e}(\varphi_G)]$. In the following, $r_{\mathbf{A}}, r_{\mathbf{B}}, r_{\mathbf{B}_t}$ respectively denote the spatial correlation coefficients obtained for the classical array, the VA prior the transform and the VA after the transform. Let's note that $r_{\mathbf{B}_t}$ depends on the source scenario $\{\theta_1, \theta_2\}$ through the transform.

Throughout this paper, a circular array with $N = 4$ antennas among which 3 are uniformly distributed around a circle of radius $0.5\lambda_0$ with λ_0 the wavelength and one central antenna is considered. Figure 1 depicts the squared spatial correlation coefficients obtained for a scenario with $M = 2$ sources of directions $\theta_1 = 180°$ and $\theta_2 = 195°$ (leading to $|r_{\mathbf{A}}(\theta_1, \theta_2)|^2 = 0.8$) with the same SNR $= 10 \log_{10}\left(\frac{\gamma_m}{\sigma^2}\right)$ of 10 dB and $K = 200$ array snapshots. Note that K is sufficiently large so that $\widehat{\mathbf{W}} \approx \mathbf{W}$ where $\mathbf{W}$ is the true whitening matrix defined from (6).

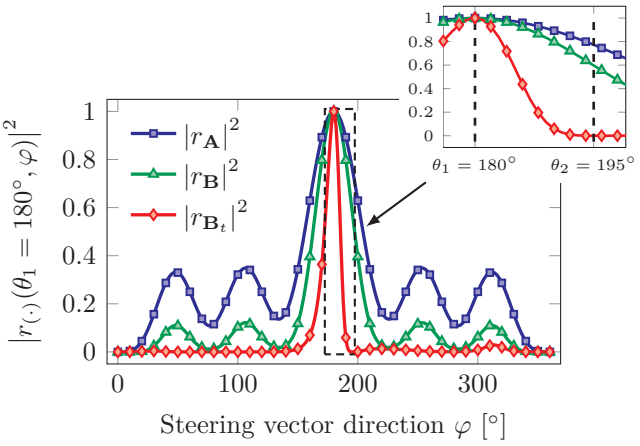

Figure 1: Squared spatial correlation as a function of the steering vector direction φ. The sources positions $\{\theta_1 = 180°, \theta_2 = 195°\}$ are represented by dashed lines.

As observed on Figure 1, the use of the VA reduces both the mainlobe width (25° for the classical array and 17° on the VA) and the sidelobes *ie.* rank-1 ambiguities since $r_{\mathbf{B}} = r_{\mathbf{A}}^2$ [11]. The two-stage transform further enhances the array resolving power by introducing a thinner mainlobe of 8° and null spatial correlation between sources directions $\{\theta_1, \theta_2\}$ as illustrated by the magnification of Figure 1. Moreover, the spatial correlation between θ_1 and many other directions is null.

The observed spatial decorrelation on Figure 1 directly modifies the eigenvalues of $\boldsymbol{\mathcal{H}}_t = \mathbf{B}_t^H(\boldsymbol{\theta})\mathbf{B}_t(\boldsymbol{\theta})$, the criterion projection onto directions Hessian matrix [19] after the two-stage transform. To quantify the eigenvalues influence on the criterion, let us consider the conditioning of the problem projection onto

directions $\{\theta_1, \theta_2\}$ defined as follows:

$$\eta_t = \frac{\lambda_{max}(\boldsymbol{\mathcal{H}}_t)}{\lambda_{min}(\boldsymbol{\mathcal{H}}_t)} \tag{29}$$

where $\lambda_{max}(\boldsymbol{\mathcal{H}}_t)$ and $\lambda_{min}(\boldsymbol{\mathcal{H}}_t)$ respectively denote the largest and smallest eigenvalues of $\boldsymbol{\mathcal{H}}_t$.

Prior the transform, the problem is badly conditioned as illustrated by the elliptical contour lines of Figure 2 obtained using the observation $\mathbf{r}$ 5. The use of the transformed observation $\mathbf{z}$ significantly improves the conditioning by reducing it of a factor 6 ($\eta = 1.1$ vs. $\eta = 6.1$) thereby leading to the almost circular contour lines shown on Figure 2.

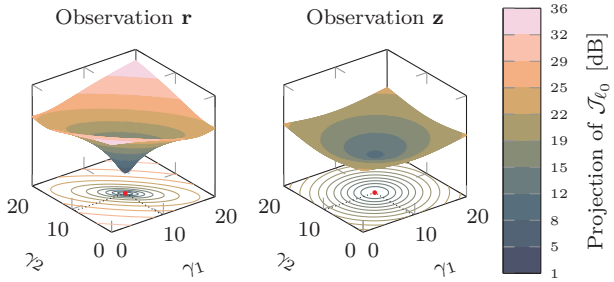

Figure 2: Projections of the sparse criterion $\mathcal{J}_{\ell_0}$ onto the sources directions $\{\theta_1 = 180°, \theta_2 = 195°\}$ before (left) and after (right) the transform for a single noise realization. Red circles and dashed lines represent the global minimum $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \hat{\gamma}_2]^T$ coordinates. Both criteria are represented using the Bartlow colormap [38]. Number of array snapshots and SNR are identical to those defined in subsection 5.1 *ie.* $K = 200$ and SNR $= 10\,$dB.

In [19], the spatial correlation coefficient expression was derived and shown to depend on the initial spatial correlation between sources $r_{\mathbf{A}}(\theta_1, \theta_2)$ and the SNR. Nevertheless, its formal connection to the problem conditioning was beyond the scope of [19]. In the following, the connection between η and $r_{\mathbf{B}_y}$ is established in the case of $M = 2$ sources.

For $M = 2$, the Hessian matrix after the transform is a 2×2 matrix with the following eigenvalues:

$$\lambda_{max/min}(\boldsymbol{\mathcal{H}}_t) = \frac{1}{2}\left(\mathrm{tr}(\boldsymbol{\mathcal{H}}_t) \pm \sqrt{\mathrm{tr}(\boldsymbol{\mathcal{H}}_t)^2 - 4\mathrm{det}(\boldsymbol{\mathcal{H}}_t)}\right) \tag{30}$$

with the $\det(\cdot)$ the determinant. Since the transformed steering matrix $\mathbf{B}_t$ has normalized columns, $\boldsymbol{\mathcal{H}}_t$ can be further simplified using eq. (28):

$$\boldsymbol{\mathcal{H}}_t = \begin{bmatrix} \mathbf{b}_t^H(\theta_1)\mathbf{b}_t(\theta_1) & \mathbf{b}_t^H(\theta_1)\mathbf{b}_t(\theta_2) \\ \mathbf{b}_t^H(\theta_2)\mathbf{b}_t(\theta_1) & \mathbf{b}_t^H(\theta_2)\mathbf{b}_t(\theta_2) \end{bmatrix} = \begin{bmatrix} 1 & r_{\mathbf{B}_t}(\theta_1,\theta_2) \\ r_{\mathbf{B}_t}(\theta_1,\theta_2) & 1 \end{bmatrix} \tag{31}$$

which immediately yields:

$$\mathrm{tr}(\boldsymbol{\mathcal{H}}_t) = 2 \quad \det(\boldsymbol{\mathcal{H}}_t) = 1 - r^2_{\mathbf{B}_t}(\theta_1,\theta_2) \tag{32}$$

Finally, substituting (32) into (30) and using the conditioning definition (29) gives:

$$\eta_t = \frac{1 + r_{\mathbf{B}_t}(\theta_1,\theta_2)}{1 - r_{\mathbf{B}_t}(\theta_1,\theta_2)} \tag{33}$$

which explicitly depends on the spatial correlation after the two-stage transform, shown to depend on both the initial spatial correlation between sources and the SNR [19]. Consequently, the problem conditioning is represented as a function of the two aforementioned quantities on Figure 3 in dB scale. Thus, the conditioning is better as $10\log_{10}(\eta)$ is close to 0 dB.

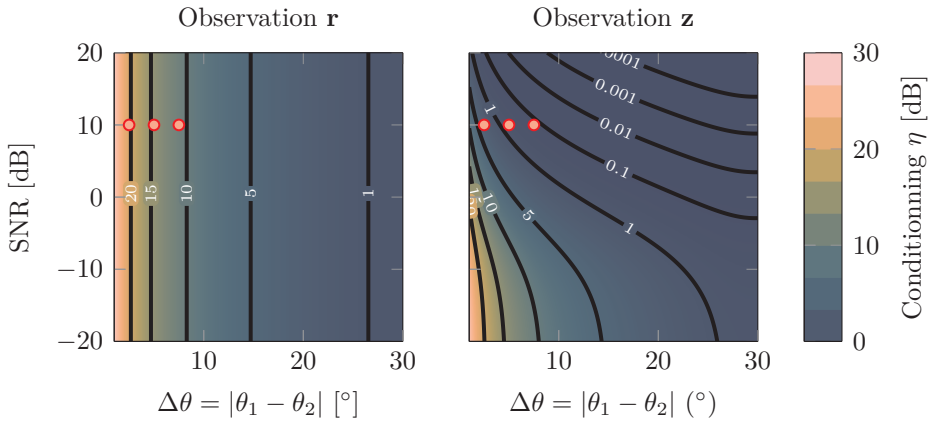

Figure 3: Problem conditioning prior (left) and after (right) the transform as a function of the initial correlation between sources and the SNR with $K = 200$ array snapshots. The red circles represent scenario parameters used in Figure 4.

Figure 3 confirms the conditioning enhancement after the transform since conditioning in significantly reduced in severe scenarios with closely spaced sources. Furthermore, for sufficient SNR, the transform leads to unitary conditioning for any angular spacing. Hence, there exists a minimal SNR that orthogonalizes the sources directions ($r_{\mathbf{B}_t}(\theta_1, \theta_2) = 0$) for a given scenario leading to $\eta = 1$.

Using the previously performed analysis on the problem conditioning, the corresponding criteria are represented for several angular spacings with fixed SNR of 10 dB on Figure 4. The improvement in conditioning leads to almost circular contour lines even if the sources are closed thus facilitating their separation and so the sparse estimator implementation.

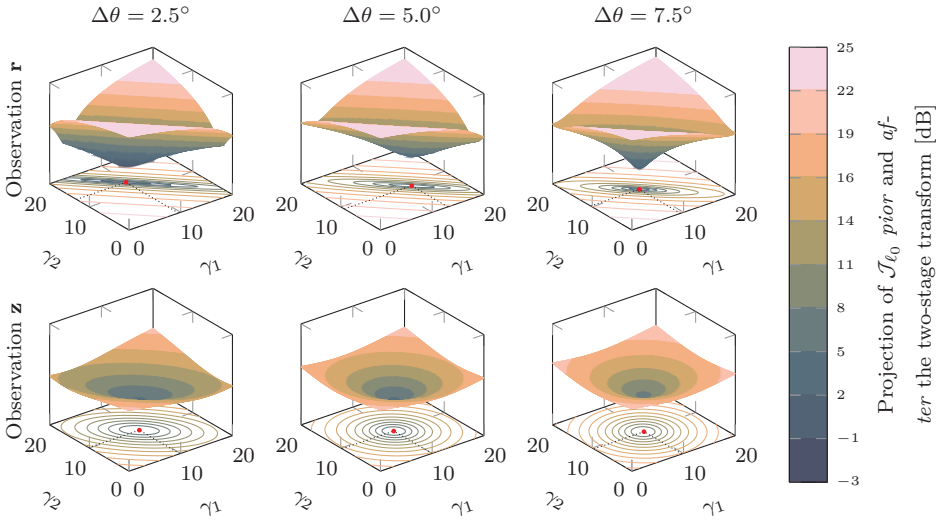

Figure 4: Influence of the angular spacing $\Delta\theta = |\theta_1 - \theta_2|$ on the sparse ℓ_0-criterion projection onto the sources directions before (top) and after (bottom) the transform. $K = 200$ array snapshots are considered together with SNR = 10 dB.

5.2. Two-stage transform effects on the minimizers of $\mathcal{J}_{CEL0}$ loss surface

The ℓ_0 penalty term in (26) enforces the solution sparsity. However, the ℓ_0-norm is a non-continuous and non-convex function that exhibits numerous local minima [39] thus grandly complexifying the optimization of the corresponding objective function. ℓ_1 convex relaxation of (26) has led to a profusion of literature since ℓ_1 and ℓ_0-formulations are equivalent under the Restricted Isometry Property (RIP) [40]. Despite its interesting theoretical guarantees, ℓ_1 relaxation is not suited to the considered application. Indeed, the RIP condition requires low correlation between pairs of dictionary vectors which is not the case with a thin grid. Furthermore, the use of ℓ_1-norm makes the penalty term sensitive to the absolute values of the sparse vector $\boldsymbol{\gamma}$ components *ie.* to the sources powers. Recently, a continuous relaxation of (26) has been proposed through the use of the CEL0 penalty [26]. The CEL0 penalty has been proved to suppress some local minima induced by the ℓ_0-norm while preserving the same global minimizer. This relaxation has been successfully applied to DoA estimation [41, 42, 18, 27].

Although $\mathcal{J}_{CEL0}$, the relaxed criterion, generally outperforms the ℓ_0-criterion, Delmer [27] experimentally demonstrated that the use of CEL0 leads to massive flat minimums where the loss surface is approximately identical in the case of closely spaced sources. These flat minimums drastically degrade the quality of the DoA estimates and cause a lack of resolution. This section investigates the effects of the transform on the minimizers of $\mathcal{J}_{CEL0}$ loss surface.

To this end, the minimum value $\mathcal{J}_{CEL0}$ is computed for all pairs of directions $\{\varphi_1, \varphi_2\}$ assuming that the sparse vector support only contains directions $\{\varphi_1, \varphi_2\}$. The corresponding surface is represented prior and after the transform on Figure 5 following the approach of [27]. $M = 2$ sources of directions $\{\theta_1 = 180°, \theta_2 = 195°\}$ with common SNR of 10 dB and $K = 200$ samples are considered for this experiment.

As observed on Figure 5, the transform suppresses the flat minimums and transforms these into sharp peaks thus improving the estimates quality.

In addition to the flat minimums suppression, the transform partially orthogonalizes the dictionary. According to Figure 1, pairs of directions $\{\theta_m, \varphi\}$

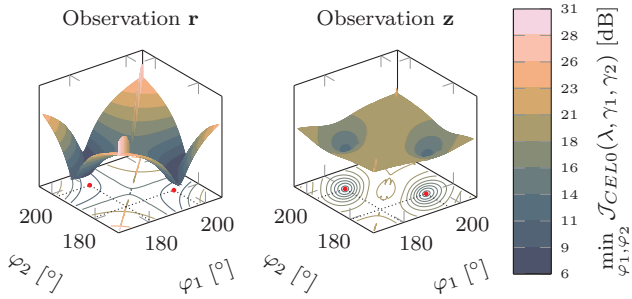

Figure 5: Loss surfaces of the CEL0 criterion prior (left) and after (right) the transform projection onto directions $\{\varphi_1, \varphi_2\}$. Red circles denote the positions of the estimated global minimums. The sources directions are $\{\theta_1 = 180°, \theta_2 = 195°\}$. Number of array snapshots and SNR are identical to those defined in subsection 5.1 *ie.* $K = 200$ and SNR $= 10\,\text{dB}$.

where θ_m is a source direction and φ is a candidate direction have low correlation. This leads to the same consequence on the sparse criterion projection onto directions $\{\theta_m, \varphi\}$ as observed on Figure 1. It is noteworthy to mention that the CEL0 penalty was initially developed assuming an orthogonal dictionary. In this particular case, the CEL0 penalty leads to a convex criterion [26]. Although, the transformed dictionary is not orthogonal, it posses many directions pairs which are orthogonal thus making it closer to the convex optimal of [26].

6. Numerical experiments

6.1. Experimental setup

To assess the performance of the proposed sparse estimator prior and after the proposed two-stage transform, both detection probability $\mathcal{P}(\theta_m)$ and Root Mean Square Error RMSE(θ_m) are measured as a function of the angular separation using 1×10^4 Monte-Carlo runs. The m-th sources is detected if the corresponding peak exists and $\left|\hat{\theta}_m - \theta_m\right| < 30°$ where $30°$ is the array half beamwidth. For the standard observation $\mathbf{r}$ (5), λ is selected using [18] whereas the results of [20] are employed for the transformed observation $\mathbf{z}$ (18). Array and all scenario parameters are identical to those defined earlier in section 5 *ie.* $K = 200$ array snapshots are considered with SNR $= 10\,\text{dB}$ for both sources.

The sparse criterion (25) is minimized using the Forward-Backward Splitting algorithm [43] applied to the CEL0 penalty. Finally, a regularly spaced grid of stepsize 1° combined an off-grid post-processing layer is employed. Specifically, this post-processing relies on the estimation of the grid-error using a first-order Taylor series expansion of the model. Further details can be found in [44].

6.2. Case $M = 2$ sources

Figure 6 summarizes the obtained performance for $M = 2$ impinging sources with respect to the angular spacing $\Delta\theta = |\theta_1 - \theta_2|$. The proposed transformed sparse estimator outperforms the classical one. After the transform, an angular spacing of only 13° is required to resolve both sources with a probability of 1. Prior the transform, this limit is increased to 20°. For this scenario, SNR = 10 dB is sufficient thus the ML on the transformed observation $\mathbf{z}$ (18) achieves the CRLB. After transform, the sparse estimator is equivalent to the ML and so achieves the CRLB as the ML does. On the contrary, the non-transformed estimator has higher RMSE than both MUSIC and the transformed estimator. This difference can be explained by the wide corridors observed on figure 5 which generate variance for the estimates.

6.3. Case $M = 3$ sources

For the case of $M = 3$ impinging sources, the directions $\{\theta_1, \theta_2, \theta_3\}$ are obtained as follows: $\theta_1 = 180°$, $\theta_2 = \theta_1 - \Delta\theta$ and $\theta_3 = \theta_1 + \Delta\theta$. All other parameters are equal to those defined in subsection 6.1. Results are represented as a function of the angular spacing $\Delta\theta$ on figure Figure 7. Note that the performances of MUSIC is not represented as it suffers from many ambiguities in the case of $M = 3$ sources with a 4-element array.

Without the two-stage transform, the estimator based on $\mathbf{r}$ (5) requires at least an angular spacing of 40° to achieve a probability of detection of 1. Furthermore, the estimator is unable to reach the CRLB. Using the proposed two-stage transform, the estimator obtained from $\mathbf{z}$ (18) achieves $\mathcal{P}(\theta_m) = 1$ for angular spacing greater than 30° thus yielding a 10° compared to the non-transformed

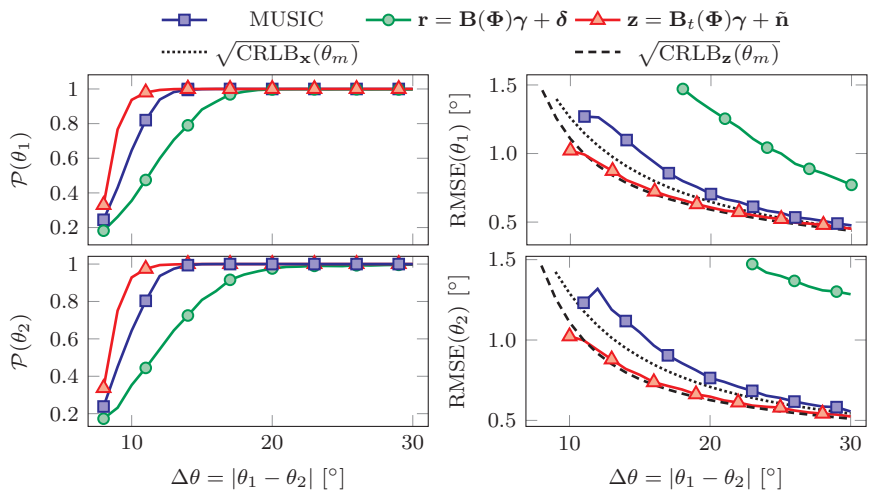

Figure 6: Probability of detection (left) and RMSE (right) for directions θ_1 (top) and θ_2 (bottom) as a function of the angular spacing. RMSE are represented for angular spacings leading to $\mathcal{P}(\theta_m) \geq 0.8$. $K = 200$ array snapshots are considered together with a common SNR of 10 dB for both sources.

estimator. The corresponding RMSE is equal to the CRLB for $\Delta\theta \geq 25°$ as the SNR is sufficient.

7. Conclusion

This paper introduces a sparse DoA estimator that implements the ML on the VCMM. Using a novel two-stage transform, equivalence with the ML is obtained. Specifically, this transform performed on the vectorized covariance matrix model converts the initially non-white and non-circular complex Gaussian noise into a real standard Gaussian noise. Thus efficient ML implementation is obtained trough sparse estimators. Additionally, the CRLB is derived to characterize the estimator asymptotic performance.

Then, the proposed transform effects on the sparse representation of the observation are analysed. The transform simplifies the sparse optimization by orthogonalizing the sources directions which drastically improves the problem conditioning. Furthermore, the transform is shown to suppress some of the

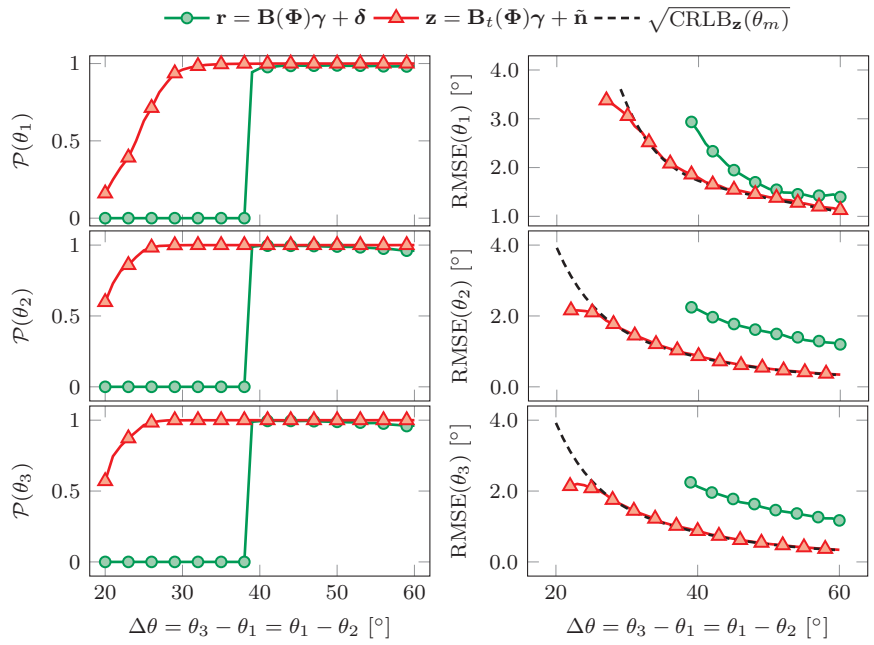

Figure 7: Probability of detection (left) and RMSE (right) for directions θ_1 (top), θ_2 (middle) and θ_3 (bottom) as a function of the angular spacing. RMSE are represented for angular spacings leading to $\mathcal{P}(\theta_m) \geq 0.8$. $K = 200$ array snapshots are considered together with a common SNR of 10 dB for both sources.

issues of the CEL0 penalty in the case of closely spaced sources. These two effects lead to consequent improvements of the sparse estimator performance in the case of closely spaced sources as confirmed by the conducted numerical simulations in the case of 2 and 3 impinging sources.

Nevertheless, a more in depth studied of the whitening transform consequences on the array's ambiguities is required as it may allow to slacken some constraints about the array geometry. Furthermore, extensions of the two-stage transform to potentially time-correlated emitters received on arrays with space and polarization diversity shall be analysed.

Acknowledgement

This research was funded by Agence Nationale de la Recherche et de la Technologie (2023/0662).

References

[1] H. Krim, M. Viberg, Two decades of array signal processing research: The parametric approach, Signal Processing Magazine, IEEE 13 (1996) 67 – 94.

[2] M. Pesavento, M. Trinh-Hoang, M. Viberg, Three more decades in array signal processing research: An optimization and structure exploitation perspective, IEEE Signal Processing Magazine 40 (4) (2023) 92–106.

[3] J. Capon, High-resolution frequency-wavenumber spectrum analysis, Proceedings of the IEEE 57 (8) (1969) 1408–1418.

[4] G. Bienvenu, L. Kopp, Optimality of high resolution array processing using the eigensystem approach, IEEE Transactions on Acoustics, Speech, and Signal Processing 31 (5) (1983) 1235–1248.

[5] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Transactions on Antennas and Propagation 34 (3) (1986) 276–280.

[6] R. Roy, A. Paulraj, T. Kailath, Estimation of signal parameters via rotational invariance techniques - esprit, in: MILCOM 1986 - IEEE Military Communications Conference: Communications-Computers: Teamed for the 90's, Vol. 3, 1986, pp. 41.6.1–41.6.5.

[7] J. F. Böhme, Estimation of spectral parameters of correlated signals in wavefields, Signal Processing 11 (4) (1986) 329–337.

[8] M. Wax, I. Ziskind, Detection of fully correlated signals by the mdl principle, 1988, pp. 2777 – 2780 vol.5.

[9] B. Ottersten, M. Viberg, P. Stoica, e. S. Nehorai, A.", J. Litva, T. J. Shepherd, Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing, Springer Berlin Heidelberg, Berlin, Heidelberg, 1993, pp. 99–151.

[10] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, On the high-snr conditional maximum-likelihood estimator full statistical characterization, IEEE Transactions on Signal Processing 54 (12) (2006) 4840–4843.

[11] P. Chevalier, L. Albera, A. Ferreol, P. Comon, On the virtual array concept for higher order array processing, IEEE Transactions on Signal Processing 53 (4) (2005) 1254–1271.

[12] P. Chevalier, A. Ferreol, L. Albera, High-resolution direction finding from higher order statistics: The $2q$-music algorithm, IEEE Transactions on Signal Processing 54 (8) (2006) 2986–2997.

[13] B. Porat, B. Friedlander, Direction finding algorithms based on high-order statistics, IEEE Transactions on Signal Processing 39 (9) (1991) 2016–2024.

[14] J. S. Picard, A. J. Weiss, Direction finding of multiple emitters by spatial sparsity and linear programming, in: 2009 9th International Symposium on Communications and Information Technology, 2009, pp. 1258–1262.

[15] Z. Yang, J. Li, P. Stoica, L. Xie, Sparse methods for direction-of-arrival estimation (2017). `arXiv:1609.09596`.

[16] D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Transactions on Signal Processing 53 (8) (2005) 3010–3022.

[17] J. Yin, T. Chen, Direction-of-arrival estimation using a sparse representation of array covariance vectors, IEEE Transactions on Signal Processing 59 (9) (2011) 4489–4493.

[18] A. Delmer, A. Ferréol, P. Larzabal, On regularization parameter for l0-sparse covariance fitting based doa estimation, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 4552–4556.

[19] T. Aussaguès, A. Ferréol, A. Delmer, P. Larzabal, Whitening effects for ml-doa estimation using a sparse representation of array covariance, in: ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025.

[20] T. Aussaguès, A. Ferréol, A. Delmer, P. Larzabal, Looking for equivalence between maximum likelihood and sparse doa estimators, in: 2024 32th European Signal Processing Conference (EUSIPCO), 2024.

[21] F. Li, R. Vaccaro, Performance degradation of doa estimators due to unknown noise fields, in: [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 1413–1416 vol.2. doi:10.1109/ICASSP.1991.150692.

[22] M. Viberg, Sensitivity of parametric direction finding to colored noise fields and undermodeling, Signal Processing 34 (2) (1993) 207–222.

[23] Q. Wu, K. M. Wong, Un-music and un-cle: an application of generalized correlation analysis to the estimation of the direction of arrival of signals in unknown correlated noise, IEEE Transactions on Signal Processing 42 (9) (1994) 2331–2343. doi:10.1109/78.317855.

[24] V. Nagesha, S. Kay, Maximum likelihood estimation for array processing in colored noise, IEEE Transactions on Signal Processing 44 (2) (1996) 169–180. doi:10.1109/78.485914.

[25] M. Pesavento, A. Gershman, Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise, IEEE Transactions on Signal Processing 49 (7) (2001) 1310–1324. doi:10.1109/78.928686.

[26] E. Soubies, L. Blanc-Féraud, G. Aubert, A Continuous Exact l0 penalty (CEL0) for least squares regularized problem, SIAM Journal on Imaging Sciences 8 (3) (2015) pp. 1607–1639 (33 pages).

[27] A. Delmer, A. Ferréol, P. Larzabal, On the complementarity of sparse l0 and cel0 regularized loss landscapes for doa estimation, Sensors 21 (18) (2021).

[28] B. Picinbono, Second-order complex random vectors and normal distributions, IEEE Transactions on Signal Processing 44 (10) (1996) 2637–2640. doi:10.1109/78.539051.

[29] H. Krim, P. Forster, J. Proakis, Operator approach to performance analysis of root-music and root-min-norm, IEEE Transactions on Signal Processing 40 (7) (1992) 1687–1696.

[30] R. J. Muirhead, Aspects of multivariate statistical theory, in: Wiley Series in Probability and Statistics, 1982.

[31] M. Mahot, F. Pascal, P. Forster, J.-P. Ovarlez, Asymptotic properties of robust complex covariance matrix estimates, IEEE Transactions on Signal Processing 61 (13) (2013) 3348–3356.

[32] J. Magnus, H. Neudecker, The commutation matrix: Some properties and applications, Annals of Statistics 7 (2) (1979) 381–394, pagination: 14.

[33] J. R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, third edition.

[34] R. Behrens, L. Scharf, Signal processing applications of oblique projection operators, IEEE Transactions on Signal Processing 42 (6) (1994) 1413–1424. doi:10.1109/78.286957.

[35] D. Slepian, Estimation of signal parameters in the presence of noise, Transactions of the IRE Professional Group on Information Theory 3 (3) (1954) 68–89.

[36] W. J. Bangs, Array processing with generalized beamformers, ph.d. dissertation.

[37] M. Nikolova, Relationship between the optimal solutions of least squares regularized with ℓ_0-norm and constrained by k-sparsity, Applied and Computational Harmonic Analysis 41 (1) (2016) 237–265, sparse Representations with Applications in Imaging Science, Data Analysis and Beyond.

[38] F. Crameri, G. E. Shephard, P. J. Heron, The misuse of colour in science communication, Nature Communications 11 (1) (2020) 5444.

[39] M. Nikolova, Description of the minimizers of least squares regularized with L0-norm. Uniqueness of the global minimizer, SIAM Journal on Imaging Sciences 6 (2) (2013) 904 – 937.

[40] E. Candes, T. Tao, Decoding by linear programming, IEEE Transactions on Information Theory 51 (12) (2005) 4203–4215. doi:10.1109/TIT.2005.858979.

[41] A. Chinatto, E. Soubies, C. Junqueira, J. a. M. T. Romano, P. Larzabal, J.-P. Barbot, L. Blanc-Féraud, ℓ_0-optimization for channel and doa sparse estimation, in: 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015, pp. 305–308.

[42] E. Soubies, A. Chinatto, P. Larzabal, J. a. M. T. Romano, L. Blanc-Féraud, Direction-of-arrival estimation through exact continuous $\ell_{2,0}$-norm relaxation, IEEE Signal Processing Letters 28 (2021) 16–20. doi:10.1109/LSP.2020.3042771.

[43] P. L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing (2010). arXiv:0912.3522.

[44] S. Bernhardt, Performances et méthodes pour l'échantillonnage comprimé : Robustesse à la méconnaissance du dictionnaire et optimisation du noyau

d'échantillonnage., Theses, Université Paris Saclay (COmUE) (Dec. 2016).
URL https://theses.hal.science/tel-01462276