

# L0-Sparse DOA Estimation of Close Sources with Modeling Errors

Alice Delmer  
Thales

SATIE, Université Paris-Saclay,  
ENS Paris-Saclay, CNRS  
Gennevilliers, France  
alice.delmer@ens-paris-saclay.fr

Anne Ferréol  
Thales

SATIE, Université Paris-Saclay,  
ENS Paris-Saclay, CNRS  
Gennevilliers, France  
anne.ferreol@satie.ens-paris-saclay.fr

Pascal Larzabal  
SATIE

Université Paris-Saclay, CNRS  
Gif-sur-Yvette, France  
pascal.larzabal@satie.ens-paris-saclay.fr

**Abstract**—In the field of array processing, Direction-Of-Arrival (DOA) estimation of close sources in the presence of modeling errors is a challenging problem. Indeed, the degradation of high-resolution methods on such scenario is well known and documented in the literature. This paper proposes an operational sparse L0-regularized method as an alternative. In sparse DOA estimation methods, the determination of the regularization parameter is a critical point, and it is generally empirically tuned. We first provide, in the presence of modeling errors, a theoretical statistical study to estimate the admissible range for this parameter in the presence of two incoming sources. For close sources, we therefore show that the admissible range is shortened. For an operational system, an off-line predetermination of the regularization parameter is required. We show that its selection is closely connected to the resolution limit for a given modeling error. Numerical simulations are presented to demonstrate the efficiency of the proposed implementation and its superiority in comparison with high-resolution methods.

**Index Terms**—DOA estimation, sparse estimation, modeling errors, close sources

## I. INTRODUCTION

In the field of high resolution Direction-Of-Arrival (DOA) estimation, subspace based methods such as MUSIC provide superior performances than other conventional techniques as beamforming or Capon's method. However, those methods are still limited for actual systems in the presence of modeling errors [1]: they have drastic limitations in the case of close sources or coherent sources. Indeed, performances depend strongly on the accuracy of the array manifold.

Sparse methods have been used for over one decade in DOA estimation [2]. The scientific community investigates these new approaches for its potential performances in challenging situations. These improvements must be established, or in other words, the scenarios for which those sparse methods are superior must be identified. Some authors highlight that those methods do not require the determination of the number of sources, or are interesting in case of a single snapshot [3]. We have recently shown in [4] that sparse methods outperform high-resolution methods in case of highly correlated sources.

In this article, we focus on the interest of sparse estimators for two close sources in presence of modeling errors. In the literature, for regularized sparse methods, the choice of the regularization parameter  $\lambda$  is often left to the user. In this

paper, thanks to a theoretical statistical analysis, we propose an automatic predetermination of  $\lambda$  based on Nikolova work [5] extended here to stochastic observations. This strategy is applied on a Single Measurement Vector (SMV) obtained from the vectorized covariance matrix assuming that the sources are independent. Required exact statistics are provided in this paper for such observation, which moreover presents great operational interests.

## II. SIGNAL MODELING AND PROBLEM FORMULATION

### A. Array signal modeling

A noisy mixture of  $M$  narrow-band sources is assumed to impinge an array of  $N$  sensors from DOAs  $\theta_m$ ,  $m = 1, \dots, M$ . Denoting  $\tilde{\mathbf{a}}(\theta) \in \mathbb{C}^N$  the steering vector of a signal source in the direction  $\theta$ , i.e. the array response for this angle, the signal at the output of the sensors is given by:

$$\mathbf{x}(t) = [x_1(t) \dots x_N(t)]^T = \sum_{m=1}^M \tilde{\mathbf{a}}(\theta_m) s_m(t) + \mathbf{n}(t), \quad (1)$$

$t = 1 \dots K$ , where  $s_m(t)$  is the complex envelope of the  $m$ th source and  $\mathbf{n}(t) \in \mathbb{C}^N$  is a white noise vector of variance  $\sigma_n^2$ .

The directions of the sources  $\theta_m$  are assumed to be in a predefined set  $\Psi = \{\psi_1, \dots, \psi_G\}$  resulting from the discretization of the field-of-view. Writing  $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}(\psi_1), \dots, \tilde{\mathbf{a}}(\psi_G)]$ , the output signal can be modeled under the sparse representation framework as:

$$\mathbf{x}(t) = \tilde{\mathbf{A}}\mathbf{s}(t) + \mathbf{n}(t) \quad (2)$$

where the vector  $\mathbf{s}(t) \in \mathbb{C}^G$  is sparse with only  $M$  non-zero entries corresponding to the sources signals  $s_m(t)$ .

But in operational systems, the steering vectors  $\tilde{\mathbf{a}}(\psi)$  of the data collection system might be different from the alleged steering vectors  $\mathbf{a}(\psi)$  used in the algorithms. Those so-called modeling errors can result from all kinds of distortions such as mutual coupling, wide-band sources etc. Let us consider additive modeling errors, accounted for example for position errors, expressed as:

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}, \quad (3)$$

with  $\mathbf{A} = [\mathbf{a}(\psi_1), \dots, \mathbf{a}(\psi_G)]$  the considered calibration matrix and  $\mathbf{E} = [\mathbf{e}(\psi_1), \dots, \mathbf{e}(\psi_G)]$  the modeling error matrix.

Let's write  $\mathbf{e}_g = \mathbf{e}(\psi_g)$ . We set  $\mathbf{a}_g^H \mathbf{a}_g = N$ . The error vector  $\mathbf{e}_g$  is assumed to be Gaussian, circular and independent on each sensor such that:  $\mathbb{E}[\mathbf{e}_g \mathbf{e}_g^H] = \sigma_e^2 \mathbf{I}_N$ ,  $\mathbb{E}[\mathbf{e}_g \mathbf{e}_g^T] = \mathbf{0}$  and  $\mathbb{E}[\mathbf{e}_g \mathbf{e}_{j \neq g}^H] = \mathbf{0}$ , where  $\mathbf{I}_N$  is the identity matrix. In practice,  $\sigma_e$  is over estimated to cover all sources of uncertainties.

### B. Vectorized covariance matrix model

In this article, the sources are supposed to be independent. Using the vectorized covariance matrix model allows to reduce the dimension of the considered observation from a  $N \times K$  matrix to an  $N^2$  vector. The effect of finite sample errors is ignored ( $K \rightarrow \infty$ ), thus the covariance matrix of the signal  $\tilde{\mathbf{R}}_{xx} = E[\mathbf{x}(t)\mathbf{x}^H(t)]$  is assumed to be known. We then have:

$$\tilde{\mathbf{R}}_{xx} = \sum_{m=1}^M \tilde{\mathbf{a}}(\theta_m) \tilde{\mathbf{a}}^H(\theta_m) \gamma_{0,m} + \sigma_n^2 \mathbf{I}_N \quad (4)$$

with  $\gamma_{0,m}$  the power of the  $m$ th source. The vectorized covariance matrix  $\tilde{\mathbf{r}} = \text{vec}(\tilde{\mathbf{R}}_{xx})$  is thus of the form

$$\tilde{\mathbf{r}} = \sum_{m=1}^M \tilde{\mathbf{a}}^*(\theta_m) \otimes \tilde{\mathbf{a}}(\theta_m) \gamma_{0,m} + \sigma_n^2 \text{vec}(\mathbf{I}_N) \quad (5)$$

$$= \sum_{m=1}^M \tilde{\mathbf{b}}(\theta_m) \gamma_{0,m} + \sigma_n^2 \text{vec}(\mathbf{I}_N), \quad (6)$$

where  $\tilde{\mathbf{b}}(\theta) = \tilde{\mathbf{a}}^*(\theta) \otimes \tilde{\mathbf{a}}(\theta)$ ,  $\mathbf{a}^*$  is the conjugate of  $\mathbf{a}$ , and  $\otimes$  is the Kronecker product.

Let's suppose that the power of the noise  $\sigma_n^2$  is known. Defining  $\tilde{\mathbf{b}}_g = \tilde{\mathbf{b}}(\psi_g)$  and  $\tilde{\mathbf{B}} = [\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_G]$ , we consider the single measurement vector (SMV) observation  $\mathbf{y} \in \mathbb{C}^{N^2}$ :

$$\mathbf{y} = \tilde{\mathbf{r}} - \sigma_n^2 \text{vec}(\mathbf{I}_N) = \tilde{\mathbf{B}} \boldsymbol{\gamma}_0, \quad (7)$$

where  $\boldsymbol{\gamma}_0 \in \mathbb{R}^G$  is  $M$ -sparse, the nonzero components represent the power of the true incoming sources and their indices indicate the corresponding directions. In presence of modeling errors, the matrix  $\tilde{\mathbf{B}}$  is unknown and the observation  $\mathbf{y}$  can thus be expressed as:

$$\mathbf{y} = \mathbf{B} \boldsymbol{\gamma}_0 + \mathbf{w}, \quad (8)$$

where  $\mathbf{B} \in \mathbb{C}^{N^2 \times G}$  is the known dictionary whose columns are  $\mathbf{b}_g = \mathbf{a}_g^* \otimes \mathbf{a}_g$ ,  $\mathbf{a}_g = \mathbf{a}(\psi_g)$ , and  $\mathbf{w}$  a noise vector. From (3),  $\mathbf{y}$  is Gaussian, with mean denoted  $\boldsymbol{\mu} \hat{=} E[\mathbf{y}]$ , and the second order statistics  $\boldsymbol{\Gamma} \hat{=} E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^H]$  and  $\mathbf{C} \hat{=} E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T]$ . Neglecting terms in  $\sigma_e^4$ , we have:

$$\boldsymbol{\mu} = \sum_{m=1}^M \mathbf{a}_m^* \otimes \mathbf{a}_m \gamma_{0,m} + \sigma_e^2 \sum_{m=1}^M \gamma_{0,m} \text{vec}(\mathbf{I}_N) \quad (9)$$

$$\boldsymbol{\Gamma} \approx \sum_{m=1}^M \gamma_{0,m}^2 (\mathbf{a}_m^* \mathbf{a}_m^T \otimes \sigma_e^2 \mathbf{I}_N + \sigma_e^2 \mathbf{I}_N \otimes \mathbf{a}_m \mathbf{a}_m^H) \quad (10)$$

$$\mathbf{C} \approx \sum_{m=1}^M \gamma_{0,m}^2 \left( \mathbf{a}_{m_1}^* \mathbf{I}_N \otimes \mathbf{a}_m, \dots, \mathbf{a}_{m_N}^* \mathbf{I}_N \otimes \mathbf{a}_m \right) \sigma_e^2 + \left[ \mathbf{a}_{m_1}^* \mathbf{I}_N \otimes \mathbf{a}_m, \dots, \mathbf{a}_{m_N}^* \mathbf{I}_N \otimes \mathbf{a}_m \right]^T \sigma_e^2 \quad (11)$$

### C. Problem formulation

The aim of sparse DOA estimation is to find the directions  $\theta_m$ : they correspond to the columns of  $\mathbf{B}$  associated with non-zero entries of  $\boldsymbol{\gamma}_0$ . We thus have to estimate  $\boldsymbol{\gamma}_0 \in \mathbb{C}^G$  from the observation  $\mathbf{y} \in \mathbb{C}^{N^2}$  (8), with  $G \gg N^2$ , using the dictionary  $\mathbf{B} \in \mathbb{C}^{N^2 \times G}$ . The problem is underdetermined, thus the knowledge that  $\boldsymbol{\gamma}_0$  is  $M$ -sparse is used, leading to the following  $\ell_0$ -constrained problem:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{B}\boldsymbol{\gamma} - \mathbf{y}\|_2^2, \quad \text{s.t.} \quad \|\boldsymbol{\gamma}\|_0 \leq M, \quad (12)$$

where the  $\ell_0$  pseudo-norm (usually referred to as  $\ell_0$ -norm) is defined as  $\|\mathbf{x}\|_0 = \text{Card}\{g \in \{1, \dots, G\} : x_g \neq 0\}$ ,  $x_g$  is the  $g$ th components of vector  $\mathbf{x}$ .

Problem (12) is NP-hard and an alternative is to find a good approximate solution of the problem under the regularized form:

$$\min_{\boldsymbol{\gamma}} \left\{ \frac{1}{2} \|\mathbf{B}\boldsymbol{\gamma} - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_0 \right\}. \quad (13)$$

The regularization parameter  $\lambda > 0$  plays a decisive role in the quality of the solutions obtained when solving this problem, but it is generally empirically tuned.

The goal is to find the best approximate solution of (12) by minimizing (13) with an automatically tuned parameter  $\lambda$  that is data independent for stochastics  $\mathbf{y}$ . Let's note that, for a given deterministic observation, Nikolova [5] determined that problems (12) and (13) are equivalent as long as  $\lambda$  belongs to a certain interval  $I$  depending on  $\mathbf{y}$ . In the next section, we extend this work to stochastic observations and apply it to the previous vectorized covariance matrix model.

## III. DETERMINATION OF AN ADMISSIBLE RANGE FOR THE REGULARIZATION PARAMETER $\lambda$

### A. Interval $I$ for the equivalence between the constrained and regularized problems

For a single observation vector  $\mathbf{y}$ , Nikolova [5] provides a condition on the regularization parameter  $\lambda$  so that the optimal solution of the regularized problem (13) is also the solution of the constrained problem (12) for a given sparsity level  $M$ .

Let's define  $c_k$  as the optimal value of the least-square residual for a vector with sparsity  $k$ :

$$c_k(\mathbf{y}) = \inf \{ \|\mathbf{B}\boldsymbol{\gamma} - \mathbf{y}\|_2^2 \mid \boldsymbol{\gamma} \in \mathbb{C}^G \text{ and } \|\boldsymbol{\gamma}\|_0 \leq k \}. \quad (14)$$

The optimal value  $c_k(\mathbf{y})$  can be computed as :

$$c_k(\mathbf{y}) = \min_{\Phi_k} (c(\mathbf{y}, \Phi_k)) \quad (15)$$

with

$$c(\mathbf{y}, \Phi_k) = \mathbf{y}^H \boldsymbol{\Pi}_n(\Phi_k) \mathbf{y}, \quad (16)$$

$$\boldsymbol{\Pi}_n(\Phi_k) = \mathbf{I} - \boldsymbol{\Pi}(\Phi_k), \quad (17)$$

$$\boldsymbol{\Pi}(\Phi_k) = \mathbf{B}(\Phi_k) [\mathbf{B}(\Phi_k)^H \mathbf{B}(\Phi_k)]^{-1} \mathbf{B}(\Phi_k)^H, \quad (18)$$

where  $\mathbf{B}(\Phi_k)$  is the submatrix of  $\mathbf{B}$  containing only the steering vectors corresponding to the  $k$  angles in the set  $\Phi_k = \{\varphi_1, \dots, \varphi_k\}$ .

The optimal sets of problems (12) and (13) coincide for a given observation  $\mathbf{y}$  if and only if  $\lambda_M^-(\mathbf{y}) < \lambda < \lambda_M^+(\mathbf{y})$  [5] with:

$$\lambda_M^-(\mathbf{y}) = \max_k \left\{ \frac{c_M(\mathbf{y}) - c_k(\mathbf{y})}{2(k-M)} \mid M < k \leq N^2 \right\}, \quad (19)$$

$$\lambda_M^+(\mathbf{y}) = \min_k \left\{ \frac{c_k(\mathbf{y}) - c_M(\mathbf{y})}{2(M-k)} \mid 0 \leq k < M \right\}. \quad (20)$$

For stochastic observations  $\mathbf{y}$ , the bounds  $\lambda_M^-$  and  $\lambda_M^+$  are also stochastic. The purpose of this paper is to theoretically study their statistics in presence of modeling errors. In this article, we will only consider the case  $M = 2$  sources. Thus we analyze the statistical behavior of  $\lambda_2^-$  and  $\lambda_2^+$ .

### B. Expression of $\lambda_2^-$ and $\lambda_2^+$ for a given observation with close sources

It is straightforward to see that the sequence  $c_k$ ,  $k = 1 \dots N^2$  verifies  $c_{N^2}(\mathbf{y}) \leq \dots \leq c_2(\mathbf{y}) \leq c_1(\mathbf{y}) \leq c_0(\mathbf{y})$ . Under weak conditions, it can be shown that:

$$\lambda_2^+(\mathbf{y}) \approx \frac{c_1(\mathbf{y}) - c_2(\mathbf{y})}{2}, \quad (21)$$

and we will assume that:

$$\lambda_2^-(\mathbf{y}) \approx \frac{c_2(\mathbf{y}) - c_3(\mathbf{y})}{2}. \quad (22)$$

Moreover, it is rather easy to derive that, for two sources incoming from directions  $\theta_1$  and  $\theta_2$ :

$$c_2(\mathbf{y}) \approx c(\mathbf{y}, \{\theta_1, \theta_2\}), \quad (23)$$

$$c_3(\mathbf{y}) \approx \min_{\psi_g \neq \theta_1, \theta_2} c(\mathbf{y}, \{\theta_1, \theta_2, \psi_g\}), \quad \psi_g \in \Psi. \quad (24)$$

Let's write:

$$Q^-(\mathbf{y}, \psi) = c(\mathbf{y}, \{\theta_1, \theta_2\}) - c(\mathbf{y}, \{\theta_1, \theta_2, \psi\}) \quad (25)$$

$$= \mathbf{y}^H (\mathbf{\Pi}_n(\{\theta_1, \theta_2\}) - \mathbf{\Pi}_n(\{\theta_1, \theta_2, \psi\})) \mathbf{y} \quad (26)$$

$$Q^+(\mathbf{y}, \psi) = c(\mathbf{y}, \psi) - c(\mathbf{y}, \{\theta_1, \theta_2\}) \quad (27)$$

$$= \mathbf{y}^H (\mathbf{\Pi}_n(\psi) - \mathbf{\Pi}_n(\{\theta_1, \theta_2\})) \mathbf{y} \quad (28)$$

Therefore, we have

$$\lambda_2^-(\mathbf{y}) \approx \frac{1}{2} \max_{\psi_g \neq \theta_1, \theta_2} Q^-(\mathbf{y}, \psi_g) \quad (29)$$

$$\lambda_2^+(\mathbf{y}) \approx \frac{1}{2} \min_{\psi_g} Q^+(\mathbf{y}, \psi_g) \quad (30)$$

with  $\psi_g \in \Psi$ . We will write  $Q_g^-(\mathbf{y}) = Q^-(\mathbf{y}, \psi_g)$  and  $Q_g^+(\mathbf{y}) = Q^+(\mathbf{y}, \psi_g)$ . In order to study the statistics of  $\lambda_2^-$  and  $\lambda_2^+$ , we will first study the statistics of  $Q_g^-$  and  $Q_g^+$ .

### C. Distributions of $\lambda_2^+$ and $\lambda_2^-$

$Q_g^-(\mathbf{y})$  and  $Q_g^+(\mathbf{y})$  are hermitian quadratic forms in non-central normal variables. An hermitian quadratic form  $Q(\mathbf{y}) =$

$\mathbf{y}^H \mathbf{P} \mathbf{y}$ , with  $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$  can be expressed as a real quadratic form  $Q(\mathbf{y}) = \mathbf{y}_R^T \mathbf{P}_R \mathbf{y}_R$  with [6]:

$$\mathbf{y}_R = [\Re(\mathbf{y})^T, \Im(\mathbf{y})^T]^T, \quad \mathbf{y}_R \sim \mathcal{N}(\boldsymbol{\mu}_R, \boldsymbol{\Gamma}_R) \quad (31)$$

$$\boldsymbol{\mu}_R = [\Re(\boldsymbol{\mu})^T, \Im(\boldsymbol{\mu})^T]^T \quad (32)$$

$$\boldsymbol{\Gamma}_R = \frac{1}{2} \begin{pmatrix} \Re(\boldsymbol{\Gamma} + \mathbf{C}) & \Im(-\boldsymbol{\Gamma} + \mathbf{C}) \\ \Im(\boldsymbol{\Gamma} + \mathbf{C}) & \Re(\boldsymbol{\Gamma} - \mathbf{C}) \end{pmatrix} \quad (33)$$

$$\mathbf{P}_R = \begin{pmatrix} (\mathbf{P} + \mathbf{P}^T)/2 & i(\mathbf{P} - \mathbf{P}^T)/2 \\ i(\mathbf{P}^T - \mathbf{P})/2 & (\mathbf{P} + \mathbf{P}^T)/2 \end{pmatrix} \quad (34)$$

Let's define  $s_k = \text{Tr}((\mathbf{P}_R \boldsymbol{\Gamma}_R)^k) + k \boldsymbol{\mu}_R^T (\mathbf{P}_R \boldsymbol{\Gamma}_R)^{k-1} \mathbf{P}_R \boldsymbol{\mu}_R$ . The mean and standard deviation of  $Q(\mathbf{y})$  are given by:  $\mu_Q = s_1$ ,  $\sigma_Q = \sqrt{2s_2}$ . Liu et al. [7] propose to approximate the distribution of a real quadratic form  $Q(\mathbf{y})$  by a non-central chi-square distribution  $\chi_l^2(\delta)$ , with  $l$  the degrees of freedom number and  $\delta$  a non-centrality parameter:

$$\Pr(Q(\mathbf{y}) > t) \approx \Pr\left(\frac{\chi_l^2(\delta) - \mu_\chi}{\sigma_\chi} > \frac{t - \mu_Q}{\sigma_Q}\right), \quad (35)$$

where  $\mu_\chi = \mathbb{E}\{\chi_l^2(\delta)\} = l + \delta$ ,  $\sigma_\chi = \sqrt{\text{var}(\chi_l^2(\delta))} = \sqrt{2a}$  and  $a = \sqrt{l + 2\delta}$ . The non-centrality parameter  $\delta$  and the degrees of freedom number  $l$  are computed so that the third-order moments of  $Q(\mathbf{y})$  and  $\chi_l^2(\delta)$  coincide, and so that the fourth-order moments are as close as possible. Defining  $b_1 = s_3/s_2^{3/2}$  and  $b_2 = s_4/s_2^2$ , it gives [7]:

- if  $b_1^2 > b_2$ ,  $\delta = b_1 a^3 - a^2$  and  $l = a^2 - 2\delta$  with  $a = 1 / (b_1 - \sqrt{b_1^2 - b_2})$ ,
- if  $b_1^2 \leq b_2$ ,  $\delta = 0$  and  $l = 1/b_1^2$ .

Knowing the distribution of the observation vector  $\mathbf{y}$ , it is therefore possible to approximate the distributions of  $Q_g^-$  and  $Q_g^+$  as non-central chi-square distributions using (35).

Moreover, the variables  $Q_g^-$ ,  $g = 1 \dots G$  (and  $Q_g^+$ ) are not independent. Let's suppose that the dependence between the variables  $Q_g^-$  for  $g = 1 \dots G$  can be modeled using a Gaussian copula with a correlation matrix  $\boldsymbol{\Sigma}$ . Higher-order moments are omitted. Let's define the random vector  $\mathbf{Q}^- = [Q_1^- \dots Q_G^-]^T$ . The approximated non-central chi-square marginal density functions of  $Q_g^-$  are denoted  $f_g^-$ , and the corresponding marginal cumulative distribution functions  $F_g^-$ . Thus, the probability density function of  $\mathbf{Q}^-$  is:

$$f_{\mathbf{Q}^-}(\mathbf{q}) = c(F_1^-(q_1), \dots, F_G^-(q_G)) f_1^-(q_1) \dots f_G^-(q_G) \quad (36)$$

where  $c$  is the probability density function of the Gaussian copula  $c(u_1, \dots, u_G) = \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2} \boldsymbol{\beta}^T (\boldsymbol{\Sigma}^{-1} - \mathbf{I}) \boldsymbol{\beta}\right)$ ,  $\boldsymbol{\beta}^T = [\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_G)]$ , and  $\Phi^{-1}$  is the inverse cumulative density function of a standard normal variable. The  $ij^{th}$  components of the correlation matrix  $\boldsymbol{\Sigma}$  are computed using the following formula for the covariance:

$$\begin{aligned} \mathbb{E}[(Q_i^- - \mathbb{E}[Q_i^-]) (Q_j^- - \mathbb{E}[Q_j^-])] &= \text{Tr}[(\mathbf{P}_i \boldsymbol{\Gamma})(\mathbf{P}_j \boldsymbol{\Gamma})] \\ &+ \text{Tr}[(\mathbf{P}_i \mathbf{C})(\mathbf{P}_j \mathbf{C})^*] + 4\boldsymbol{\mu}^H \mathbf{P}_i \boldsymbol{\Gamma} \mathbf{P}_j \boldsymbol{\mu} \end{aligned} \quad (37)$$

where  $Q_k^-(\mathbf{y}) = \mathbf{y}^H \mathbf{P}_k \mathbf{y}$  with  $\mathbf{P}_k = \mathbf{\Pi}_n(\{\theta_1, \theta_2\}) - \mathbf{\Pi}_n(\{\theta_1, \theta_2, \psi_k\})$  for  $k = i, j$ , and  $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$ .

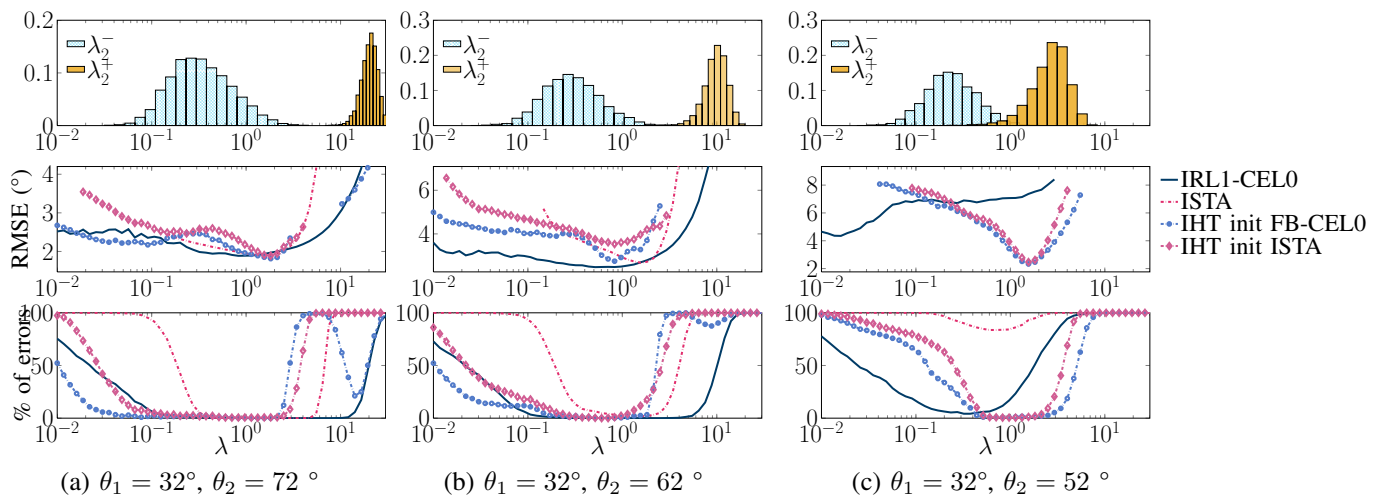


Fig. 1. Top: empirical histograms of  $\lambda_2^-$  and  $\lambda_2^+$  for  $\sigma_e = 0.2$ . Middle: RMSE (in degrees) for the different algorithms as a function of the regularization parameter  $\lambda$ . Bottom: percentage of errors for the same algorithms. Beamwidth =  $44^\circ$ .

The distribution function of  $\lambda_2^-$  can then be computed as:

$$F_{\lambda_2^-}(t) = \Pr\left(\frac{1}{2} Q^-(\mathbf{y}, \psi_g) < t, \forall \psi_g \in \{\Psi \setminus \theta_1, \theta_2\}\right). \quad (38)$$

However, in practice, we will use an empirical distribution function  $\hat{F}_{\lambda_2^-}$  obtained from the maximum of samples of  $\mathbf{Q}^-$  drawn using (36). The approach is similar for  $\mathbf{Q}^+$  and  $\lambda_2^+$ . Note that obtaining an empirical distribution directly from (14) (19) and (20) is combinatorial: the previous statistical study allows to drastically reduce the complexity.

These statistics allow, with high probability, a data independent determination of an admissible interval  $I = [F_{\lambda_2^-}^{-1}(1 - \epsilon^-); F_{\lambda_2^-}^{-1}(\epsilon^+)]$ .  $\epsilon^-$  and  $\epsilon^+$  are small values adjusting the confidence interval, as long as there is no overlapping between both statistics.

#### IV. SIMULATION RESULTS

We consider an Uniform Circular Array (UCA) with  $N = 7$  antennas and radius  $d = \lambda_0/2$ , where  $\lambda_0$  is the wavelength. UCA are well known for its  $\theta$  invariant performances: for  $M = 2$  sources, we are thus interested in the behavior depending on the spacing between the two sources.

Iterative Hard Thresholding (IHT) [8] is the most commonly used algorithm to solve (13), but might only converge to a local minima of the cost function. Results of IHT alone are not presented here as in most cases they are not conclusive: having a correct initialization is fundamental. The recently developed Continuous Exact- $\ell_0$  penalty (CELO) [10] aims to improve the global convergence of  $\ell_0$ -based algorithms and can be used with a Forward-Backward (FB-CELO) or an IRL1 algorithm (IRL1-CELO) to minimize the  $\ell_0$  criterion [11]. In this article, we propose to combine CELO and IHT to use advantages of both methods. We show that the result of FB-CELO can be used as an initialization to IHT algorithm. We will use Iterative Soft Thresholding Algorithm (ISTA) [9], the low-computational method for the minimization of the

convex  $\ell_1$ -regularized criterion, as a comparison to  $\ell_0$  based methods, although sufficient conditions to recover a sparse signal using  $\ell_1$  instead of  $\ell_0$  are usually too restrictive in practical applications. Note that in general, our theoretical analysis on the regularization parameter  $\lambda$  does not apply for  $\ell_1$  minimization. However, for the settings used in this article, we will see that the same parameter can be used.

##### A. Interval for stochastic observations

Empirical distributions of  $\lambda_2^-$  and  $\lambda_2^+$  are computed as in (29) and (30) using 3000 numerical samples of  $\mathbf{Q}_2^-$  and  $\mathbf{Q}_2^+$  generated from (36) for  $\sigma_e = 0.2$ . Histograms are represented on Fig. 1 (top) for three different spacings inside the main beamwidth ( $44^\circ$ ). Performances of the algorithms are presented over 1000 independent trials as a function of the parameter  $\lambda$  in terms of Root-Mean Square Error (RMSE) between the estimated directions and the true DOAs (middle), and the percentage of errors (bottom). The estimation is considered erroneous if only one source (or none) is detected or if the bias between an estimated direction and the true DOA is superior to half a beamwidth. The trial is then excluded from the RMSE computation. Whenever the percentage of error is superior to 80%, the RMSE is not represented.

For high spacing ( $40^\circ$ ), an admissible interval for  $\lambda$  is clearly visible for all algorithms on Fig. 1(a): the parameter  $\lambda$  shall not be over the lower limit of  $\lambda_2^+$ , but can possibly take a value under the maximum of  $\lambda_2^-$ . Histograms of  $\lambda_2^-$  and  $\lambda_2^+$  are closer when the sources are closer. When the histograms start to overlap (just before  $20^\circ$  of spacing), only IHT with a suitable initialization can provide an interval where the percentage of errors is almost (but never) null. Results of FB-CELO are not represented: for high spacing, it acts like IRL1-CELO, and for lower spacing like ISTA. However, it allows consistent results in all regimes as an initialization to IHT. On the contrary, IRL1-CELO can not be used to initialize IHT: the convergence of the algorithm is already too strong.

### B. Data independent selection of the regularization parameter

Based on the previous analysis, we propose in this paper to select  $\lambda$  at the intersection between  $\hat{F}_{\lambda_2^-}^{-1}(0.98)$  and  $\hat{F}_{\lambda_2^+}^{-1}(0.02)$ , computed as a function of the spacing  $|\theta_2 - \theta_1|$ , as represented on Fig. 2. For  $\sigma_e = 0.2$ , the intersection occurs at  $\lambda = 0.92$  for  $|\theta_2 - \theta_1| \approx 20^\circ$ . We can infer that this choice is attached to the resolution limit.

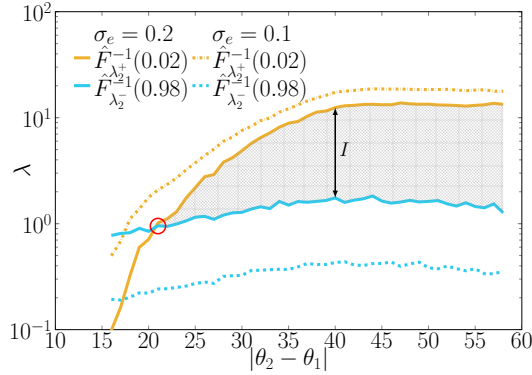


Fig. 2. Upper limit of the histogram of  $\lambda_2^-$  and lower limit of the histogram of  $\lambda_2^+$  as a function of the difference  $|\theta_2 - \theta_1|$  for  $\sigma_e = 0.2$  and  $\sigma_e = 0.1$ . Intersection define the Pareto front.

The value of  $\sigma_e$  is never known exactly: in order to establish the performances of an operational system, an upper bound can be estimated. If  $\sigma_e$  turns out to be lower than the estimated value, the proposed choice of  $\lambda$  will still be valid, as shown on Fig. 2, as it remains between the bounds of  $\lambda_2^-$  and  $\lambda_2^+$ .

### C. Performances for the selected $\lambda$

RMSE and percentage of errors for the selected  $\lambda$  are compared to the performances of MUSIC. Results are represented on Fig. 3 for  $\sigma_e = 0.2$ , and on Fig. 4 for  $\sigma_e = 0.1$ . IHT algorithm initialized by ISTA or FB-CEL0 outperforms the other methods in terms of resolution limit, but FB-CEL0 initialization provides better performances regarding RMSE. Resolution limit with the proposed method is  $20^\circ$  for  $\sigma_e = 0.2$  and  $17^\circ$  for  $\sigma_e = 0.1$ , as expected from Fig. 2 for  $\lambda = 0.92$ , whereas it is only  $35^\circ$  and  $26^\circ$  with MUSIC. Compared methods have analogue performances for highly spaced sources.

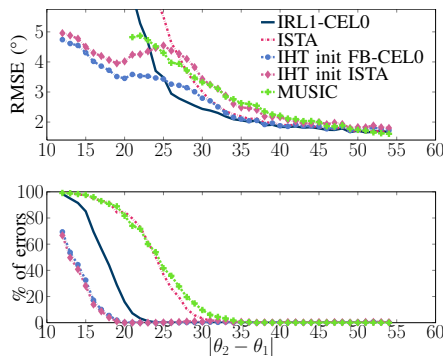


Fig. 3. RMSE in degrees between the estimated and true DOAs, and percentage of errors as a function of  $|\theta_2 - \theta_1|$ , for  $\sigma_e = 0.2$  and  $\lambda = 0.92$ .

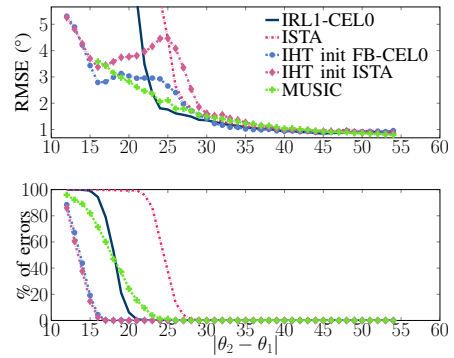


Fig. 4. RMSE in degrees between the estimated and true DOAs, and percentage of errors as a function of  $|\theta_2 - \theta_1|$ , for  $\sigma_e = 0.1$  and  $\lambda = 0.92$ .

## V. CONCLUSION

This work focuses on  $\ell_0$ -sparse DOA estimation for close sources in presence of modeling errors under a regularized formulation. A data independent determination of an admissible range for the regularization parameter  $\lambda$  is provided. An operational strategy consists in selecting the value of  $\lambda$  on the Pareto front. It is applied on the vectorized covariance matrix for which the statistics have been obtained in presence of modeling errors. For close sources, simulations highlight the superiority of the proposed  $\ell_0$  regularization implementation methods in comparison with  $\ell_1$  and MUSIC methods.

## REFERENCES

- [1] B. Friedlander, "A Sensitivity Analysis of the MUSIC Algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, 1990.
- [2] J. J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Transactions on Signal Processing*, 2001.
- [3] Z. Yang, J. Li, P. Stoica, and L. Xie, "Chapter 11 - sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, pp. 509 – 581.
- [4] A. Delmer, A. Ferréol, and P. Larzabal, "Selection of regularization parameter in sparse inverse problems for doa estimation," *to be published in : Journal of Physics: Conference Series*, 2019.
- [5] M. Nikolova, "Relationship between the optimal solutions of least squares regularized with L0-norm and constrained by k-sparsity," *Applied and Comp. Harmonic Analysis*, vol. 41, no. 1, pp. 237–265, 2016.
- [6] A. A. Mohsenipour and S. B. Provost, "On the Distribution of Quadratic Expressions in Various Types of Random Vectors," Ph.D. dissertation, 2012.
- [7] H. Liu, Y. Tang, and H. H. Zhang, "A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables," *Computational Statistics and Data Analysis*, vol. 53, no. 4, pp. 853–856, feb 2009.
- [8] T. Blumensath and M. E. Davies, "Iterative Thresholding for Sparse Approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, 2008.
- [9] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [10] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A Continuous Exact L0 penalty (CEL0) for least squares regularized problem," *SIAM J. Imaging Sciences*, vol. 8, 2015.
- [11] S. Gazagnes, E. Soubies, and L. Blanc-Féraud, "High density molecule localization for super-resolution microscopy using cel0 based sparse approximation," in *14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 28–31.