



## **Selection of regularization parameter in sparse inverse problems for DOA estimation**

Alice Delmer, Anne Ferréol, Pascal Larzabal

### **► To cite this version:**

Alice Delmer, Anne Ferréol, Pascal Larzabal. Selection of regularization parameter in sparse inverse problems for DOA estimation. *Journal of Physics: Conference Series*, 2020, 1476, pp.012001. <10.1088/1742-6596/1476/1/012001>. <hal-03153852>

**HAL Id: hal-03153852**

**<https://hal.science/hal-03153852v1>**

Submitted on 26 Feb 2021

**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.



HAL Authorization

# Selection of regularization parameter in sparse inverse problems for DOA estimation

Alice Delmer<sup>1,2</sup>, Anne Ferréol<sup>1,2</sup> and Pascal Larzabal<sup>2</sup>

<sup>1</sup> Thales, 4 avenue des Louvresses, Gennevilliers, France

<sup>2</sup> SATIE, CNRS, ENS Paris-Saclay, Université Paris-Saclay, 61 avenue du Président Wilson, Cachan, France

E-mail: `name.surname@satie.ens-cachan.fr`

**Abstract.** This article deals with the selection of the regularization parameter of a sparse  $\ell_0$  regularized criterion for DOA (Direction-of-Arrival) estimation. This parameter is generally empirically obtained. In this paper, we establish a theoretical range for this parameter allowing a correct estimation for a single observation. Considering that an observation is the eigenvector associated with the maximum eigenvalue of the covariance matrix, we extend the analysis to multiple observations. Numerical results support our theoretical investigations.

## 1. Introduction

Conventional methods for Direction-Of-Arrival (DOA) estimation such as beamforming and Capon's method [1] are looking for peaks in the energy spectrum in the field of view. To improve their performances, DOA estimation schemes can be obtained in the framework of inverse problems. For that, subspace based method (high-resolution) such as MUSIC [2] have been introduced. These former methods suffer from well known limitations in case of coherent sources or a limited number of available snapshots. Moreover, as encountered in inverse problems, modeling errors can degrade drastically their performances [3]. As an alternative to this drawback, a calibration table, which is the matrix of the array recorded responses in various directions of the field of view, is usually used in operational systems. The framework of sparse representation fits naturally with the use of a calibration table used in place of an analytic array manifold far from the true one. The advantages of sparse methods have been widely studied in the last decade in order to improve performances of DOA estimation in challenging situations. An overview is provided in [4].

This article focuses on the minimization of the  $\ell_0$ -regularized criterion. The main issue of recent works on  $\ell_0$  implementation is the choice of the regularization parameter  $\lambda$ , which is traditionally empirically tuned. The aim of this work is to study the influence of  $\lambda$  in order to propose a strategy for its determination.

## 2. Bearing estimation as a sparse inverse problem

### 2.1. Signal modeling

Let us consider  $M$  narrowband sources impinging on an array consisting of  $N$  antennas from angles  $\theta_m$ ,  $m = 1 \dots M$ . The corresponding arriving signals are denoted  $\tilde{s}_m(t)$  and the sensor

outputs  $z_n(t)$ ,  $n = 1 \dots N$ . The array output is modeled by :

$$\mathbf{z}(t) = [z_1(t) \dots z_N(t)]^T = \sum_{m=1}^M \mathbf{a}(\tilde{\theta}_m) \tilde{s}_m(t) + \mathbf{n}(t), \quad (1)$$

where  $\mathbf{a}(\tilde{\theta}_m) \in \mathbb{C}^N$ , called steering vector, is the array response for angle  $\tilde{\theta}_m$ , with  $\|\mathbf{a}(\tilde{\theta}_m)\|_2^2 = N$ , and  $\mathbf{n}(t) \in \mathbb{C}^N$  is an additive noise. Denoting the true directions of arrival  $\tilde{\Theta} = \{\tilde{\theta}_1, \dots, \tilde{\theta}_M\}$ , let us define the mixing matrix of the sources

$$\tilde{\mathbf{A}} = \mathbf{A}(\tilde{\Theta}) = [\mathbf{a}(\tilde{\theta}_1) \dots \mathbf{a}(\tilde{\theta}_M)]. \quad (2)$$

We then have :

$$\mathbf{z}(t) = \tilde{\mathbf{A}} \tilde{\mathbf{s}}(t) + \mathbf{n}(t), \quad \text{with } \tilde{\mathbf{s}}(t) = [\tilde{s}_1(t), \dots, \tilde{s}_M(t)]^T. \quad (3)$$

## 2.2. Sparse representation

We are working with a calibration table  $\mathbf{A} \in \mathbb{C}^{N \times G}$  containing the array responses for discrete directions  $\Theta = \{\theta_1, \dots, \theta_G\}$ , with  $G \gg M$  :

$$\mathbf{A} = \mathbf{A}(\Theta) = [\mathbf{a}(\theta_1) \dots \mathbf{a}(\theta_G)]. \quad (4)$$

In this paper, the true directions of the sources are supposed to be on the predefined grid used for sparse representation. The mixing matrix  $\tilde{\mathbf{A}}$  is thus a submatrix of the calibration table and  $\tilde{\Theta} \subset \Theta$ .

Assuming that the number of directions  $G$  in the calibration table verifies  $G \gg N$ , (3) can be rewritten as an underdetermine system :

$$\mathbf{z}(t) = \mathbf{A} \mathbf{s}(t) + \mathbf{n}(t), \quad (5)$$

where the vector  $\mathbf{s}(t)$  of dimension  $G \times 1$  is sparse with  $M$  non-zero entries corresponding to the sources signals  $\tilde{s}_m(t) : s_{i_m}(t) = \tilde{s}_m(t)$ . In sparse DOA estimation, the purpose is to recover the support  $I_M = \{i_m, m = 1 \dots M\}$ , where associated columns in  $\mathbf{A}$  correspond to the sources steering vectors  $\mathbf{a}(\tilde{\theta}_m)$ , and indicate then the sources directions  $\{\tilde{\theta}_1, \dots, \tilde{\theta}_M\}$ .

In a more general way, let's consider the sparse model of a vector observation :

$$\mathbf{y} = \mathbf{A} \mathbf{x}_0 + \mathbf{n}_0, \quad \text{with } \mathbf{A} \mathbf{x}_0 = \tilde{\mathbf{A}} \tilde{\mathbf{x}}_0, \quad (6)$$

where  $\mathbf{x}_0$  is the true value, and the  $M \times 1$  vector  $\tilde{\mathbf{x}}_0$  is composed of all nonzero entries of  $\mathbf{x}_0$ . The Single Measurement Vector (SMV) in (5) is directly of the form of the general model (6), in which  $\mathbf{y} = \mathbf{z}(t_k)$  and  $\mathbf{x}_0 = \mathbf{s}(t_k)$ .

In order to improve accuracy and robustness, multiple snapshots can be used. We define the Multi Measurement Vectors (MMV)  $\mathbf{Z}_K$  as  $\mathbf{Z}_K = [\mathbf{z}(t_1), \dots, \mathbf{z}(t_K)]$  and then :

$$\mathbf{Z}_K = \mathbf{A} \mathbf{S}_K + \mathbf{N}_K, \quad (7)$$

with  $\mathbf{S}_K = [\mathbf{s}(t_1), \dots, \mathbf{s}(t_K)]$  and  $\mathbf{N}_K = [\mathbf{n}(t_1), \dots, \mathbf{n}(t_K)]$ . Matrix  $\mathbf{S}_K$  is row-sparse : sparsity lies, in the case of DOA estimation, in space and not in time. Although it could be possible to seek  $\mathbf{S}_K$  from measurements  $\mathbf{Z}_K$ , Malioutov et al. [5] propose to use Singular Value Decomposition (SVD) in order to reduce the dimension and thus the computational cost. SVD can be employed either on the measurement matrix  $\mathbf{Z}_K$  or on the covariance matrix  $\mathbf{R}_{zz} = E[\mathbf{z}(t)\mathbf{z}^H(t)]$ .

Eigenvectors  $\mathbf{e}_m$ ,  $m = 1, \dots, M$ , corresponding to the  $M$  maximal eigenvalues, are a sparse combination of columns of  $\mathbf{A}$  :

$$\mathbf{e}_m = \sum_{g=1}^G \mathbf{a}(\theta_g) x_{g,m} = \sum_{i=1}^M \mathbf{a}(\tilde{\theta}_i) \tilde{x}_{i,m}. \quad (8)$$

In this paper, we will consider an observation which is the eigenvector corresponding to the highest eigenvalue :

$$\mathbf{e}_1 = \mathbf{A} \mathbf{x}_{0_e} + \mathbf{n}_{0_e}, \quad (9)$$

which is of the form (6) with  $\mathbf{y} = \mathbf{e}_1$ . Lets note that the first eigenvector is easily obtained thanks to the power method.

### 2.3. Sparse estimation schemes

In the following, we consider the model of (6) with a single observation vector. We want thus to recover the vector  $\mathbf{x}_0 \in \mathbb{C}^G$  from observation  $\mathbf{y} \in \mathbb{C}^N$ . The problem is underdetermined ( $G \gg N$ ), so the minimum least-square solution is generally not unique. In order to obtain the unique solution, the sparsity of vector  $\mathbf{x}_0$  must be exploited. The sparsity of a vector  $\mathbf{x}$  is measured by the so-called  $\ell_0$ -norm defined as  $\|\mathbf{x}\|_0 = \text{Card} \{g \in \{1, \dots, G\} : x_g \neq 0\}$ , where  $\text{Card} \{\cdot\}$  is the cardinal of  $\{\cdot\}$ .

To uniquely recover  $\mathbf{x}_0$  from (6), the following constrained minimization problem must be solved :

$$\min_{\mathbf{x} \in \mathbb{C}^G} \|\mathbf{A} \mathbf{x} - \mathbf{y}\|_2^2, \quad \text{subject to} \quad \|\mathbf{x}\|_0 \leq k, \quad (10)$$

with  $k$  a sparsity level. In DOA estimation, according to (5)(6), this sparsity should correspond to the number of sources  $M$ . This problem is known to be NP-hard. It is also possible to formulate (10) as the minimization of the  $\ell_0$ -norm with a constraint on the square error, related to the noise. As finding the optimal solution requires an exhaustive search, different sub-optimal approaches can be used. The main class of algorithms solving the constrained  $\ell_0$ -problem are *greedy* algorithms, such as Matching Pursuit [6] or Orthogonal Matching Pursuit [7]. Those algorithms have known limitations : the support of the solution is gradually increased by one element, and a false detection can never be deleted in a further step.

It is then worth to formulate the problem as a regularized problem, in which we seek to minimize a cost function  $J(\lambda, \cdot)$  :

$$\min_{\mathbf{x} \in \mathbb{C}^G} \left\{ J(\lambda, \mathbf{x}) = \frac{1}{2} \|\mathbf{A} \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_0 \right\}, \quad (11)$$

Regularization parameter  $\lambda > 0$  aims to balance the squared error and the sparsity in the cost function. The minimum of the cost function and the optimal solution  $\hat{\mathbf{x}}_0$  depends on the parameter  $\lambda$ . For a fixed  $\lambda$ , we define  $J_\lambda$  as :

$$J_\lambda : \begin{array}{ccc} \mathbb{C}^G & \longrightarrow & \mathbb{R} \\ \mathbf{x} & \longmapsto & J(\lambda, \mathbf{x}) \end{array} \quad (12)$$

This cost function can be minimized by the Iterative Hard Thresholding (IHT) algorithm [8] [9]. However, only a local convergence is guaranteed.

An alternative is to relax the  $\ell_0$ -norm by the  $\ell_1$ -norm, which is continuous and convex. Both constraint problem (10) and regularized problem (11) can then be solved more easily : the  $\ell_1$ -constrained problem, known under the name *Basis Pursuit* or *Basis Pursuit Denoising (BPDN)* [10], can be solved via Linear Programming (LP), and the  $\ell_1$ -regularized problem,

known as LASSO (*Least Absolute Shrinkage and Selection Operator*) can be solved by the Iterative Soft-Thresholding Algorithm (ISTA) [11]. Moreover, when the restricted isometry property (RIP) is satisfied, it is equivalent to  $\ell_0$ -norm resolution [12].

However, as the calibration table is a discretized recording of the array responses for close DOAs, the columns of  $\mathbf{A}$  might be highly correlated and thus the RIP is generally not verified. Consequently, pseudo  $\ell_0$ -norm must be necessary used in such challenging problems. Recent papers use  $\ell_0$ -norm approximations in order to improve accuracy or global convergence of  $\ell_0$  based algorithms. In this way, we can cite for example the works of [13] or [14] in which  $\ell_0$ -norm is approximate by gaussian functions. In this paper, we choose to work with the recently developed CEL0 penalty [15]. Indeed, their authors have proven that CEL0-regularized criterion exhibits less local minima than  $\ell_0$ -criterion and in [16] they showed that CEL0 also improves statistical performances. The CEL0-regularized criterion can be minimized by a *Forward-Backward* proximal algorithm ; we will note it FB-CEL0. The difficulty is now to estimate the regularization parameter.

### 3. Selection of the regularization parameter

#### 3.1. Theoretical background for a single observation

Our approach is based on the work of [17], which established connections between the solutions of the  $\ell_0$ -regularized problem (11) and the  $k$ -constrained problem (10). The so-obtained relationship between the sets of global minimizers of the two problems allows to give a range on the regularization parameter  $\lambda$  so that the optimal solution of the minimization of the cost function  $J_\lambda$  corresponds to our scenario with  $M$  sources.

Let us define the optimal value  $c_k$  and the set of optimal solutions  $\hat{X}_c^k$  of (10) as :

$$c_k = \inf \{ \|\mathbf{Ax} - \mathbf{y}\|_2^2 \mid \mathbf{x} \in \mathbb{C}^G \text{ and } \|\mathbf{x}\|_0 \leq k \} \quad (13)$$

$$\hat{X}_c^k = \{ \mathbf{x} \in \mathbb{C}^G \text{ and } \|\mathbf{x}\|_0 \leq k \mid \|\mathbf{Ax} - \mathbf{y}\|_2^2 = c_k \} \quad (14)$$

Lets consider an example of DOA estimation with  $M$  impinging sources. In presence of  $M$  sources, for  $k = M$ , the set  $\hat{X}_c^M$  must contain only one element in order to estimate  $\mathbf{x}_0$  without ambiguities. Denote by  $L$  the minimal integer so that the optimal value  $c_L$  is null; it satisfies  $L \leq N$ . It is straightforward to see that the sequence  $c_k$ ,  $k = 1 \dots L$  verifies  $c_L \leq \dots \leq c_{M+1} \leq c_M \leq c_{M-1} \leq \dots \leq c_0$ , and that  $\hat{\mathbf{x}}_k \in \hat{X}_c^k$  for  $k \leq L$  satisfies  $\|\hat{\mathbf{x}}_k\|_0 = k$ .

For a fixed  $\lambda$ , let us now define the set of optimal solutions  $\hat{X}_r^\lambda$  of (11) as :

$$\hat{X}_r^\lambda = \{ \mathbf{x} \in \mathbb{C}^G \mid J_\lambda(\mathbf{x}) = r_\lambda \}, \text{ with } r_\lambda = \inf \{ J_\lambda(\mathbf{x}) \mid \mathbf{x} \in \mathbb{C}^G \}. \quad (15)$$

We are looking for the values of the parameter  $\lambda$  so that optimal solutions of problems (10) and (11) coincide, i.e.  $\hat{X}_r^\lambda = \hat{X}_c^k$ . For any  $k \leq L$ ,  $\forall \hat{\mathbf{x}}_k \in \hat{X}_c^k$ , one has  $J(\lambda, \hat{\mathbf{x}}_k) = \frac{1}{2}c_k + \lambda k$ . For all  $\mathbf{x}_k$  such that  $\|\mathbf{x}_k\|_0 = k$ , it holds that  $J_\lambda(\hat{\mathbf{x}}_k) \leq J_\lambda(\mathbf{x}_k)$ .

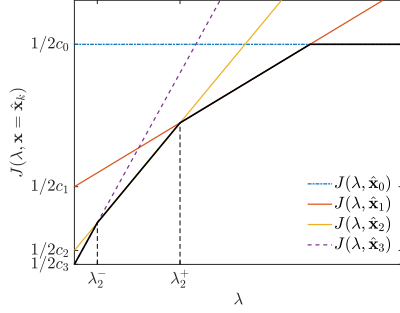
For  $\hat{\mathbf{x}}_M \in \hat{X}_c^M$ , we wish to find  $\lambda$  such that  $\forall k \neq M$  and  $\hat{\mathbf{x}}_k \in \hat{X}_c^k$ ,  $J(\lambda, \hat{\mathbf{x}}_M) \leq J(\lambda, \hat{\mathbf{x}}_k)$ , i.e :

$$\frac{1}{2}c_M + \lambda M < \frac{1}{2}c_k + \lambda k. \quad (16)$$

From (16), we deduce that  $\lambda$  needs to satisfy  $\lambda_M^- < \lambda < \lambda_M^+$ , with :

$$\lambda_M^+ = \min_k \left\{ \frac{c_k - c_M}{2(M - k)} \mid 0 \leq k < M \right\} \text{ and } \lambda_M^- = \max_k \left\{ \frac{c_M - c_k}{2(k - M)} \mid M < k \leq L \right\} \quad (17)$$

Fig. 1 represents a typical example of the affine functions  $J(\lambda, \mathbf{x} = \hat{\mathbf{x}}_k)$  as a function of  $\lambda$ , for  $k = 0, \dots, 3$ . In this work, we set  $M = 2$  ; consequently, there is a range  $[\lambda_2^-, \lambda_2^+]$  so that  $\forall \lambda \in [\lambda_2^-, \lambda_2^+]$ ,  $\min_\lambda J(\lambda, \hat{\mathbf{x}}_k) = J(\lambda, \hat{\mathbf{x}}_2)$ . Note that the affine property of functions  $J(\lambda, \mathbf{x})$  are exploited in [18] to propose another approach for  $\ell_0$ -regularized problems.

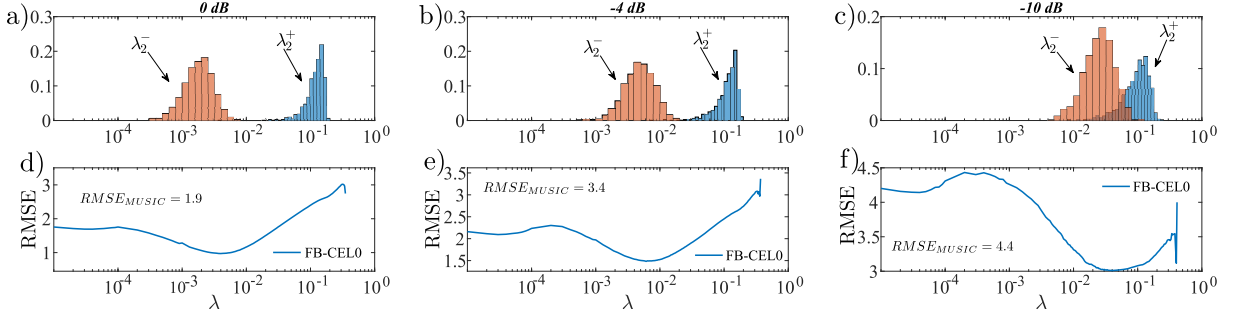


**Figure 1.** Functions  $J(\lambda, \mathbf{x})$  as a function of  $\lambda$ , evaluated in  $\mathbf{x} = \hat{\mathbf{x}}_k$ ,  $k = 0 \dots 3$ .  $\lambda_2^-$  and  $\lambda_2^+$  delimit the range where the minimum of  $J(\lambda, \mathbf{x})$  (in black) corresponds to  $J(\lambda, \hat{\mathbf{x}}_2)$ .

### 3.2. Numerical statistical analysis of $\lambda_2^-$ and $\lambda_2^+$

We consider a Uniform Circular Array (UCA) composed of  $N = 7$  antennas, placed along a circle of radius  $r = 0.8\lambda_0$ , with  $\lambda_0$  the wavelength. The 3dB beamwidth is  $26^\circ$ . We will consider  $M = 2$  impinging sources, and  $K = 100$  snapshots to compute the eigenvector used as the observation. We study a range of Signal-to-Noise Ratio (SNR) between 0 and -15dB, and wish to improve performances of MUSIC in this context.

**3.2.1. Statistical dispersion for one DOAs scenario** Values  $\lambda_2^-$  and  $\lambda_2^+$  depend a priori on the SNR, and on the linear combination of the two steering vectors in the estimated eigenvector. This combination is different for each trial. For the DOAs scenario with  $\tilde{\theta}_1 = 40^\circ$  and  $\tilde{\theta}_2 = 120^\circ$ , statistics of  $\lambda_2^-$  and  $\lambda_2^+$  are showed in figure 2 for (a) SNR = 0dB, (b) SNR = -4dB, and (c) SNR = -10dB.

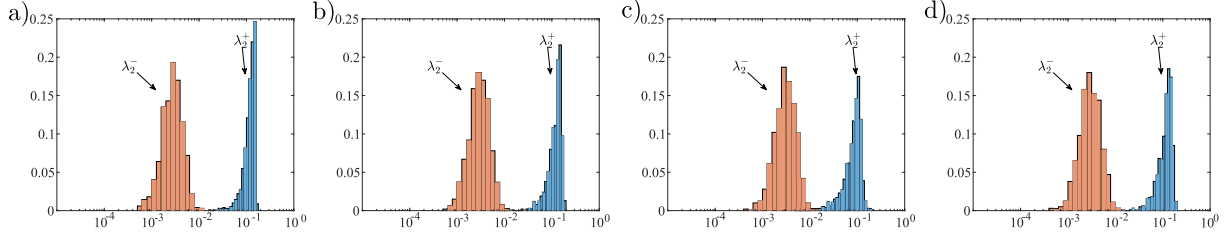


**Figure 2.** Statistics of  $\lambda_2^-$  and  $\lambda_2^+$  for (a) SNR = 0dB, (b) SNR = -4dB, and (c) SNR = -10dB. Root-Mean-Square Error (RMSE), in degrees, as a function of  $\lambda$  for the same SNRs (d,e,f).

**3.2.2. Pseudo-stability regarding the DOAs** It can be easily proved that for an uniform circular array, the Cramer-Rao Bound is independent of the DOA [19]. In our two sources scenario, this result holds approximatively as soon as the source separation is greater than the beamwidth. Consequently, we infer that the statistics of  $\lambda_2^-$  and  $\lambda_2^+$  will be  $\Delta\theta$  independent. This stability is clearly showed for different values of  $\Delta\theta = |\tilde{\theta}_1 - \tilde{\theta}_2|$  in Fig. 3. This property is very interesting in the sense that an estimation of the noise power will be sufficient to define the regularization parameter range whatever the (unknown) scenario.

## 4. Simulation results for DOA estimation

We consider the previous scenario with  $\tilde{\theta}_1 = 40^\circ$ ,  $\tilde{\theta}_2 = 120^\circ$ . The sources are additionally temporally coherent, with a correlation coefficient  $c = 0.9$ . If not stated otherwise, we performed

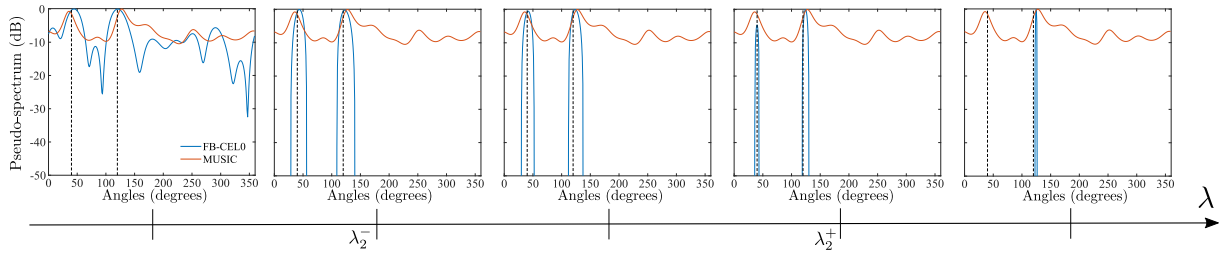


**Figure 3.** Statistics of  $\lambda_2^-$  and  $\lambda_2^+$  for different DOAs at SNR = -2dB : (a)  $\Delta\theta = 40^\circ$ , (b)  $\Delta\theta = 80^\circ$ , (c)  $\Delta\theta = 120^\circ$ , (d)  $\Delta\theta = 160^\circ$ .

1000 Monte-Carlo trials to recover the statistics.

#### 4.1. For a given observation

An example of result obtained with FB-CEL0 is showed in Fig. 4. As expected from our theoretical investigation, if  $\lambda < \lambda_2^-$ , the resolution capability is limited. If  $\lambda_2^- \leq \lambda \leq \lambda_2^+$ , we can observe two thin peaks corresponding to the true DOAs. If  $\lambda > \lambda_2^+$ , the algorithm might no more exhibit two peaks.



**Figure 4.** An example of results of FB-CEL0 as a function of  $\lambda$ , compared with MUSIC pseudo spectrum.

#### 4.2. For a stochastic observation

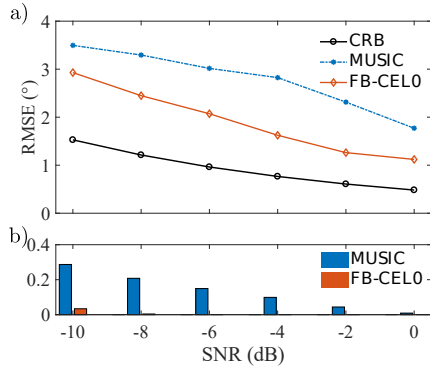
Fig. 2 shows the Root-Mean-Square Error (RMSE) in degrees for (d) SNR = 0dB, (e) SNR = -4dB, and (f) SNR = -10dB. For  $\lambda > \max(\lambda_2^+)$ , the algorithm fails and the solution is the null vector. If  $\lambda$  is too small, the algorithm behaves as the least-squares. For each SNR, although statistics of  $\lambda_2^-$  and  $\lambda_2^+$  can overlap (depending on the SNR), there is a range for  $\lambda$  giving better results than MUSIC. The theoretical estimation of this optimal range is an ongoing work.

#### 4.3. Performances analysis and comparisons

Performances of FB-CEL0 are compared to the Cramer-Rao Bound and MUSIC in Fig. 5 (a). FB-CEL0 outperforms MUSIC performances. In Fig 5 (b), we have reported the probability of mis-localization (outside an interval corresponding to one beamwidth and centered on the true direction). We observe that the robustness of MUSIC degrades drastically in comparison with FB-CEL0. For those simulations, we choose  $\lambda$  as the median of the statistics of  $\lambda_2^-$  for each SNR.

## 5. Conclusion

This work has focused on the selection of the  $\ell_0$  regularization parameter for a sparse bearing estimation. In the case of a single observation, we obtained a theoretical range for  $\lambda$  and



**Figure 5.** (a) Root-Mean-Square Error of MUSIC (\*) and FB-CEL0 ( $\diamond$ ), compared with the Cramer-Rao Bound ( $\circ$ ) for different Signal-to-Noise Ratio (SNR). (b) Probability of mis-localization for MUSIC and FB-CEL0.

validated it on simulations. In the case of multiple observations and for an UCA, we have shown the independence of  $\lambda$  with the DOAs. This important property allows a selection of  $\lambda$  depending only on the SNR, which can be easily estimated online.

## References

- [1] Capon J 1969 High-Resolution frequency-wavenumber spectrum analysis *Proc. of the IEEE* **57**
- [2] Bienvenu G and Kopp L 1983 Optimality of high resolution array processing using the eigensystem approach *IEEE Trans. on Acoustics, Speech, and Signal Processing*
- [3] Friedlander B 1990 Sensitivity analysis of the MUSIC algorithm *IEEE Trans. on Acoustics, Speech, and Signal Processing* **38** 1740–51
- [4] Yang Z, Li J, Stoica P and Xie L 2017 Sparse methods for direction-of-arrival estimation (Academic Press Library in Signal Processing vol 7) pp 509–581
- [5] Malioutov D, Çetin M and Willsky A S 2005 A sparse signal reconstruction perspective for source localization with sensor arrays *IEEE Trans. on Signal Processing* **53**
- [6] Mallat S G and Zhang Z 1993 Matching pursuits with time-frequency dictionaries *IEEE Trans. on Signal Processing* **41**
- [7] Pati Y C, Rezaiifar R and Krishnaprasad P S 1993 Orthogonal matching pursuit : recursive function approximation with applications to wavelet decomposition *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers* **1** pp 40–44
- [8] Blumensath T and Davies M E 2008 Iterative thresholding for sparse approximations *J. of Fourier Analysis and Applications*
- [9] — 2010 Normalized iterative hard thresholding: guaranteed stability and performance *IEEE J. on Selected Topics in Signal Processing*
- [10] Chen S S, Donoho D L and Saunders M A 1998 Atomic decomposition by basis pursuit *SIAM J. on Scientific Computing*
- [11] Donoho D L 1995 De-noising by soft-thresholding *IEEE Transactions on Information Theory* **41**
- [12] Candes E J 2008 The restricted isometry property and its implications for compressed sensing *Comptes Rendus Mathématiques* **346** 9
- [13] Mohimani H, Babaie-Zadeh M and Jutten C 2009 A fast approach for overcomplete sparse decomposition based on smoothed L0-norm *IEEE Trans. on Signal Processing*
- [14] Hyder M M and Mahata K 2010 Direction-of-Arrival estimation using a mixed L2,0 norm approximation *IEEE Trans. on Signal Processing* **58** 9
- [15] Soubies E, Blanc-Féraud L and Aubert G 2015 A Continuous Exact L0 penalty (CEL0) for least squares regularized problem *SIAM J. Imaging Sciences* **8**
- [16] Chinatto A, Soubies E, Junqueira C, Romano J M, Larzabal P, Barbot J P and Blanc-Féraud L 2015 0-optimization for channel and DOA sparse estimation *IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*
- [17] Nikolova M 2016 Relationship between the optimal solutions of least squares regularized with l0-norm and constrained by k-sparsity *Applied and Computational Harmonic Analysis* **41** pp 237–265
- [18] Duan J, Soussen C, Brie D and Idier J 2009 A continuation approach to estimate a solution path of mixed L2-L0 minimization problems
- [19] Stoica P and Nehorai A 1989 MUSIC, Maximum Likelihood, and Cramer-Rao Bound *IEEE Trans. on Acoustics, Speech, and Signal Processing* **37**