



**HAL**  
open science

# On the Complementarity of Sparse L0 and CEL0 Regularized Loss Landscapes for DOA Estimation

Alice Delmer, Anne Ferréol, Pascal Larzabal

► **To cite this version:**

Alice Delmer, Anne Ferréol, Pascal Larzabal. On the Complementarity of Sparse L0 and CEL0 Regularized Loss Landscapes for DOA Estimation. *Sensors*, 2021, 21 (18), pp.6081. 10.3390/s21186081 . hal-04451148

**HAL Id: hal-04451148**

**<https://hal.science/hal-04451148>**

Submitted on 5 Apr 2024

**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.

Communication

# On the Complementarity of Sparse L0 and CEL0 Regularized Loss Landscapes 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, 92230 Gennevilliers, France; anne.ferreol@satie.ens-cachan.fr<sup>2</sup> SATIE, ENS Paris-Saclay, Université Paris-Saclay, CNRS, 91190 Gif-sur-Yvette, France; pascal.larzabal@satie.ens-cachan.fr

\* Correspondence: alice.delmer@ens-paris-saclay.fr

**Abstract:** L0 sparse methods are not widespread in Direction-Of-Arrival (DOA) estimation yet, despite their potential superiority over classical methods in difficult scenarios. This comes from the difficulties encountered for global optimization on hill-climbing error surfaces. In this paper, we explore the loss landscapes of L0 and Continuous Exact L0 (CEL0) regularized problems in order to design a new optimization scheme. As expected, we observe that the recently introduced CEL0 penalty leads to an error surface with less local minima than the L0 one. This property explains the good behavior of the CEL0-regularized sparse DOA estimation problem for well-separated sources. Unfortunately, CEL0-regularized landscape enlarges L0-basins in the middle of close sources, and CEL0 methods are thus unable to resolve two close sources. Consequently, we propose to alternate between both error surfaces to increase the probability of reaching the global solution. Experiments show that the proposed approach offers better performance than existing ones, and particularly an enhanced resolution limit.

**Keywords:** direction-of-arrival; sparse modeling; L0 regularization; nonconvex regularization; loss landscapes



**Citation:** Delmer, A.; Ferréol, A.; Larzabal, P. On the Complementarity of Sparse L0 and CEL0 Regularized Loss Landscapes for DOA Estimation. *Sensors* **2021**, *21*, 6081. <https://doi.org/10.3390/s21186081>

Academic Editor: Sisi Zlatanova

Received: 21 July 2021

Accepted: 7 September 2021

Published: 10 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The study of Direction-Of-Arrival (DOA) estimation has a long history in signal processing. Conventional methods [1] such as beamforming or Capon's method are still the subject of numerous works, e.g., [2]. However, they present degraded performance in the presence of multiple close sources. Subspace-based methods such as Multiple Signal Classification (MUSIC) have been introduced to improve the resolution limit for multiple sources. Unfortunately, these methods fail in presence of correlated sources [3]. They also often require a priori knowledge of the number of sources and need a sufficient number of snapshots. Sparse DOA estimation has received much attention in the last decade due to its potential performance in such scenarios [4–10].

Sparsity naturally arises in DOA estimation when considering a discretization of the field-of-view in numerous candidate angles of arrival on a grid. The aim is to estimate a vector  $\gamma$  whose dimension covers the whole grid and whose only the very few entries corresponding to sources DOA are non-zero.

The purpose of sparse estimation, under the Single Measurement Vector (SMV) framework, is to retrieve the sparse vector  $\gamma \in \mathbb{C}^G$  from the noisy measurement  $\mathbf{y} = \mathbf{B}\gamma + \mathbf{w}$ , with  $\mathbf{y} \in \mathbb{C}^{N^2}$ ,  $N^2 \ll G$ , knowing a dictionary  $\mathbf{B}$ . The dictionary  $\mathbf{B}$  depends on the array's responses for the different angles of arrival candidates. It can be formulated as the following regularized problem:

$$\min_{\gamma} \left\{ J_{\ell_0}(\lambda, \gamma) = \frac{1}{2} \|\mathbf{B}\gamma - \mathbf{y}\|_2^2 + \lambda \|\gamma\|_0 \right\}, \quad (1)$$

where the so-called  $\ell_0$ -norm is defined as  $\|\gamma\|_0 = \text{Card}\{g \in \{1, \dots, G\} : \gamma_g \neq 0\}$ ,  $\gamma_g$  being the  $g$ -th component of vector  $\gamma$ . The  $\ell_0$ -norm is the natural measure of sparsity: it counts the number of non zero components of the vector. The regularization parameter  $\lambda$  aims to balance the relative importance between the data fidelity term  $1/2\|\mathbf{B}\gamma - \mathbf{y}\|_2^2$  and the  $\ell_0$ -norm enforcing sparsity of the solution. The sparse estimation problem can also be formulated as a constraint problem. The relationship between the two  $\ell_0$ -problems has been studied in [11]. Based on this study, recent theoretical results [12,13] have been provided for an off-line selection of  $\lambda$  so that the  $\ell_0$ -problems are equivalent. The regularization parameter is here chosen in accordance with those results.

The  $\ell_0$ -minimization problem is known to be NP-hard: its resolution usually requires an exhaustive search. The use of the very recently proposed global optimization method [14] is limited to small size problems and is thus unadapted here because of its huge computational cost. So far, many suboptimal methods have therefore been proposed, as the well-known Iterative Hard Thresholding (IHT) algorithm. IHT is a proximal gradient descent algorithm: it iteratively produces estimates  $\hat{\gamma}^{(i)}$  so that the cost function  $J_{\ell_0}$  decreases, starting from an initial point  $\hat{\gamma}^{(0)}$ . However, the  $\ell_0$ -regularized error surface  $J_{\ell_0}$  exhibits numerous local minima, and convergence is only proved to a stationary point. Convex relaxation of (1) by the  $\ell_1$ -norm is also a popular alternative. However, conditions [15] under which the sparse vector can be reliably recovered are usually too restrictive for practical applications as in DOA estimation.

More recently, minimization of a regularized criterion using nonsmooth nonconvex but continuous penalties has drawn considerable attention [16], and it has been shown in many applications that it can yield significantly better performance than with using the  $\ell_1$ -norm [17]. Such penalties include  $\ell_q$ -norms ( $0 < q < 1$ ), Smoothly Clipped Absolute Deviation (SCAD), and Minimax Concave Penalty (MCP) [18]. The Continuous Exact  $\ell_0$  (CEL0) penalty [19] corresponds to the limit case of MCP for the SMV framework. CEL0 is shown to suppress some local minima of  $J_{\ell_0}$  while preserving the global one. The CEL0-regularized cost function is:

$$J_{\text{CEL0}}(\lambda, \gamma) = \frac{1}{2}\|\mathbf{B}\gamma - \mathbf{y}\|_2^2 + \Phi_{\text{CEL0}}(\lambda, \gamma), \quad (2)$$

with

$$\Phi_{\text{CEL0}}(\lambda, \gamma) = \sum_{i \in \mathbb{I}_G} \phi(\lambda, \alpha_i, \gamma_i) \quad (3)$$

$$\phi(\lambda, \alpha_i, \gamma_i) = \lambda - \frac{(|\gamma_i| - \sqrt{2\lambda\alpha_i})^2}{2\alpha_i} \mathbb{1}_{\{|\gamma_i| \leq \sqrt{2\lambda\alpha_i}\}} \quad (4)$$

and  $\alpha_i = 1/\|\mathbf{B}_{\cdot,i}\|_2^2$ ,  $\mathbf{B}_{\cdot,i}$  being the  $i$ -th column of matrix  $\mathbf{B}$ .  $\mathbb{1}$  is the indicator function whose value is one if the given condition is respected and zero otherwise. Despite its promising interest, we have shown in [12] that traditional suboptimal optimization schemes of CEL0-regularized functional as Iterative Reweighted  $\ell_1$  (IRL1) or Forward Backward (FB) are unable to resolve close sources.

The aim of this paper is to investigate the properties of  $J_{\ell_0}$  and  $J_{\text{CEL0}}$  loss surfaces in order to propose a sparse optimization strategy to resolve close sources. The goal is to improve the resolution limit of both MUSIC method, limited for low Signal-to-Noise Ratios, and existing sparse methods. The proposed approach follows an iterative scheme that requires little computational cost. It shows good performance for close sources and does not require any particular initialization.

Outline of the paper: we first explain the model and the sparse DOA estimation problem in Section 2. In Section 3, we compare  $J_{\ell_0}$  and  $J_{\text{CELO}}$  loss surfaces in the context of multi-source DOA estimation: an in-deep analysis of the minimizers is provided for two close sources. Based on this analysis, Section 4 presents the proposed optimization scheme, whose originality is to alternate between both loss surfaces. Numerical simulations of Section 5 finally show the validity and advantages of our approach.

Notations: Upper-case and lower-case boldface letters denote matrices and vectors, respectively.  $\cdot^*$  denotes the conjugate,  $\cdot^T$  the transpose and  $\cdot^H$  the conjugate transpose of a vector or matrix.  $x_i$  is the  $i^{\text{th}}$  component of vector  $\mathbf{x}$ , and  $\omega_i$  the  $i^{\text{th}}$  component of the set  $\omega$ . Given a matrix  $\mathbf{X}$ , the  $i^{\text{th}}$  column is denoted  $\mathbf{X}_{\cdot i}$ . Considering a matrix  $\mathbf{X}$  of dimension  $N \times G$ ,  $\mathbf{X}_\omega$  is the submatrix of  $\mathbf{X}$  containing the columns indicated by the set  $\omega \subseteq \mathbb{I}_G$ , where  $\mathbb{I}_G = \{1, \dots, G\}$  is the ordered index set. Similarly,  $\mathbf{x}_\omega$  is the subvector of  $\mathbf{x}$  defined as:  $\mathbf{x}_\omega = (x_{\omega_1}, \dots, x_{\omega_{\#\omega}})^T$ , with  $\#\omega$  the number of elements in  $\omega$ .

## 2. Sparse DOA Estimation Problem

### 2.1. On-Grid Array Signal Modeling

Consider  $M$  far field narrow band sources impinging an array of  $N$  antennas from angles  $\tilde{\theta}_m$ ,  $m = 1 \dots M$ . For a single snapshot at time  $t$ , the output array signal  $\mathbf{x}(t) \in \mathbb{C}^N$  is expressed as  $\mathbf{x}(t) = [x_1(t) \dots x_N(t)]^T = \sum_{m=1}^M \mathbf{a}(\tilde{\theta}_m) \tilde{s}_m(t) + \mathbf{n}(t)$ , where  $\mathbf{a}(\tilde{\theta}_m)$  is the steering vector (or array response) for the direction  $\tilde{\theta}_m$ ,  $\tilde{s}_m(t)$  the complex envelope of the signal of the  $m$ th source, and  $\mathbf{n}(t) \in \mathbb{C}^N$  a white gaussian noise vector of covariance  $E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2 \mathbf{I}_N$ , where  $\mathbf{I}_N$  is the  $N \times N$  identity matrix. Let us suppose the directions of the sources are part of a predefined set  $\Theta = \{\theta_1, \dots, \theta_G\}$  resulting from the discretization of the field-of-view, with  $G \gg N$ : for all arrival angles  $\tilde{\theta}_m$ ,  $m \in [1, \dots, M]$ , there exists  $g \in [1, \dots, G]$  such that  $\tilde{\theta}_m = \theta_g$ . This assumption is often considered in operational systems to measure the calibration table  $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_G)]$  containing the array responses for the angles in  $\Theta$ . Considering this calibration table, the measurement  $\mathbf{x}(t)$  can be expressed as:

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

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

Under the assumption that the sources are uncorrelated, it is interesting to use the vectorized covariance matrix in algorithms: it allows us to consider the contribution of multiple snapshots thus increasing the accuracy, without increasing the computational cost by much. It also has the advantage of being a SMV model, thus all associated methods can be used, and it additionally gives the possibility of estimating more sources than the number of sensors.

### 2.2. Vectorized Covariance Matrix Model

Considering uncorrelated sources, the covariance matrix  $\mathbf{R}_{xx} \triangleq E[\mathbf{x}(t)\mathbf{x}^H(t)]$  is given by:

$$\mathbf{R}_{xx} = \sum_{m=1}^M \mathbf{a}(\tilde{\theta}_m) \mathbf{a}^H(\tilde{\theta}_m) \tilde{\gamma}_m + \sigma_n^2 \mathbf{I}_N, \quad (6)$$

with  $\tilde{\gamma}_m$  the power of the  $m$ th source and  $\mathbf{I}_N$  the square identity matrix of dimension  $N$ . The vectorized covariance matrix, noted as  $\mathbf{r} = \text{vec}(\mathbf{R}_{xx})$ , is the vector obtained from the concatenation of the columns of  $\mathbf{R}_{xx}$ . It can be expressed as

$$\mathbf{r} = \sum_{m=1}^M \mathbf{b}(\tilde{\theta}_m) \tilde{\gamma}_m + \sigma_n^2 \text{vec}(\mathbf{I}_N) \quad (7)$$

with  $\mathbf{b} = \mathbf{a}^* \otimes \mathbf{a}$ ,  $\mathbf{a}^*$  is the conjugate of  $\mathbf{a}$ , and  $\otimes$  is the Kronecker product. Considering a dictionary  $\mathbf{B} = [\mathbf{b}(\theta_1), \dots, \mathbf{b}(\theta_G)]$  computed from the calibration matrix  $\mathbf{A}$ , we have

$\mathbf{r} = \mathbf{B}\boldsymbol{\gamma} + \sigma_n^2 \text{vec}(\mathbf{I}_N)$ . Considering  $K$  finite samples, the covariance matrix is estimated by  $\hat{\mathbf{R}}_{xx} = 1/K \sum_{k=1}^K \mathbf{x}(t_k) \mathbf{x}^H(t_k)$ , and we denote  $\hat{\mathbf{r}}$  the associated estimated vectorized covariance matrix. Let us suppose that the power of the noise is known. We consider the noisy observation vector  $\mathbf{y} \in \mathbb{C}^{N^2}$  computed as:

$$\mathbf{y} = \hat{\mathbf{r}} - \sigma_n^2 \text{vec}(\mathbf{I}_N) = \mathbf{B}\boldsymbol{\gamma} + \mathbf{w}, \quad (8)$$

with  $\mathbf{B} \in \mathbb{C}^{N^2 \times G}$ . The noise vector  $\mathbf{w}$  results from the estimation of  $\mathbf{R}_{xx}$  with a finite number of snapshots. The power vector  $\boldsymbol{\gamma} \in \mathbb{C}^G$  is sparse and the indices of non-zero components indicate the directions of the sources. The aim of sparse DOA estimation is to retrieve the indices of non-zero components of vector  $\boldsymbol{\gamma}$  in Equation (8), through the resolution of the problem given by Equation (1).

### 3. Description and Numerical Investigations of the Minimizers of $J_{\ell_0}$ and $J_{\text{CEL0}}$

It is known that there are numerous local minima in  $J_{\ell_0}$ , which complicates the minimization of the criterion. Initialization of iterative descent algorithms is particularly delicate, as also highlighted in the DOA estimation literature. In [12], we have successfully used CEL0 penalty for DOA estimation of well-separated sources. Those good results do not seem to be transposed to the case of close sources. It is important now to analyze more deeply the minimizers of  $\ell_0$  and CEL0 penalized problems in order to propose an optimization scheme able to resolve closer sources.

#### 3.1. Simulation Setup

Although our approach is array and scenario independent, we illustrate it in this paper with the following setup. We consider an Uniform Circular Array (UCA) with  $N = 7$  antennas and radius  $d = \lambda_0/2$ , where  $\lambda_0$  is the wavelength. This array could allow for two-dimensions direction-of-arrival estimation, but we limit ourselves to azimuth estimation. The  $-3$  dB beamwidth of this array is  $40^\circ$ . UCAs are well known for their  $\theta$  invariant performance. We study the case of  $M = 2$  incoming sources located at  $\tilde{\theta}_1 = 32^\circ$  and a varying  $\tilde{\theta}_2$ . The number of snapshots is fixed to  $K = 50$ . In this part, the received signal is noiseless. The field-of-view is the range  $[0, 360]^\circ$  with a grid spacing of  $0.5^\circ$  ( $G = 720$ ). The mutual coherence, corresponding to the maximum absolute correlation between two columns of the dictionary, is in this case close to 1:  $\ell_1$  methods are thus ineffective.

#### 3.2. Minimizers of $J_{\ell_0}$

Let us define  $\mathbb{I}_G = \{1, \dots, G\}$  the ordered index set. For a given observation  $\mathbf{y} \in \mathbb{C}^G$  and a set  $\omega \subseteq \mathbb{I}_G$ , we define the constrained problem ( $\mathcal{C}_\omega$ ) as follows:

$$(\mathcal{C}_\omega) : \min_{\boldsymbol{\gamma}} \|\mathbf{B}\boldsymbol{\gamma} - \mathbf{y}\|_2^2, \quad \text{s. t. } \gamma_i = 0, \quad \forall i \in \mathbb{I}_G \setminus \omega \quad (9)$$

where  $\gamma_i$  is the  $i$ th component of vector  $\boldsymbol{\gamma}$ . Let us note  $\hat{\boldsymbol{\gamma}}_\omega$  the subvector of  $\hat{\boldsymbol{\gamma}}$  composed only with the terms indicated by  $\omega$ :  $\hat{\boldsymbol{\gamma}}_\omega = \left( \hat{\gamma}_{\omega_1}, \dots, \hat{\gamma}_{\omega_{\#\omega}} \right)^T$ , with  $\#\omega$  the number of elements in  $\omega$ . The constraint ensures the solutions  $\hat{\boldsymbol{\gamma}}$  are sparse vectors whose indices of all non zero entries are in  $\omega$ , i.e., all components whose indices are not in  $\omega$  are null. We can then write  $\hat{\boldsymbol{\gamma}} = Z_p(\hat{\boldsymbol{\gamma}}_\omega)$ , where  $Z_p$  is a zero padding operator in  $\mathbb{I}_G$ :

$$\hat{\gamma}_i = \begin{cases} 0 & \text{if } i \notin \omega \\ \hat{\gamma}_{\omega_k} & \text{for the unique } k \text{ such that } \omega_k = i. \end{cases} \quad (10)$$

For any  $\omega \subseteq \mathbb{I}_G$ ,  $\hat{\gamma} \in \mathbb{C}^G$  solves  $(\mathcal{C}_\omega)$  if and only if  $\hat{\gamma}_\omega \in \mathbb{C}^{\#\omega}$  solves  $\mathbf{B}_\omega^H \mathbf{B}_\omega \mathbf{x} = \mathbf{B}_\omega^H \mathbf{y}$  and  $\hat{\gamma} = Z_p(\hat{\gamma}_\omega)$ .  $\mathbf{B}_\omega$  is the submatrix of  $\mathbf{B}$  composed only with the columns whose indices are in  $\omega$ .

There are strong connections between the minimizers of the constrained problem  $(\mathcal{C}_\omega)$  and the minimizers of the regularized criterion  $J_{\ell_0}(\lambda, \cdot)$  that we want to analyze. For  $\mathbf{y} \in \mathbb{C}^N$ , given a set  $\omega \subseteq \mathbb{I}_G$ , let  $\hat{\gamma}$  solve problem  $(\mathcal{C}_\omega)$ . Then for any  $\lambda$ ,  $J_{\ell_0}(\lambda, \cdot)$  reaches a (local) minimum at  $\hat{\gamma}$  (Proposition 2.3 [20]). Conversely, for  $\mathbf{y} \in \mathbb{C}^N$  and  $\lambda > 0$ , let  $J_{\ell_0}(\lambda, \cdot)$  have a (local) minimum at  $\hat{\gamma}$ . Then  $\hat{\gamma}$  solves  $(\mathcal{C}_{\hat{\omega}})$  for  $\hat{\omega} = \text{supp}(\hat{\gamma})$  (Lemma 2.4 [20]). Moreover, the (local) minimum that  $J_{\ell_0}(\lambda, \cdot)$  has at  $\hat{\gamma}$  is strict iff  $\text{rank}(\mathbf{B}_{\hat{\omega}}) = \#\hat{\omega}$  (Theorem 3.2 [20]). There is thus a large number of local minima: the number of supports  $\omega \subseteq \mathbb{I}_G$  such that  $\text{rank}(\mathbf{B}_\omega) = \#\omega$  that lead to strict local minima is upper bounded by  $\sum_{k=0}^{N_2} \binom{G}{k}$ .

In [20], it is shown that under mild conditions,  $J_{\ell_0}(\lambda, \cdot)$  have a unique strict global minimizer. It is common knowledge that the optimal solution of the regularized problem given by Equation (1) depends on the regularization parameter  $\lambda$ , which balances the relative importance between data fidelity and sparsity. In most papers, this parameter is empirically tuned. In previous works [12,13], we proposed a theoretical analysis for an off-line selection of  $\lambda$ . In the sequel of this paper,  $\lambda$  will be selected in an appropriate interval  $I$  as defined in [12,13].

Figure 1a–d represents projections of  $J_{\ell_0}$  for the scenario described above, for  $\tilde{\theta}_1 = 32^\circ$  and  $\tilde{\theta}_2 = 62^\circ$  in the noiseless case. Iso-levels of  $J_{\ell_0}$  are reported for a vector  $\gamma$  having at most two non-zero components:  $\gamma = Z_p(\gamma_\omega)$  with  $\#\omega = 2$ . Those components correspond to fixed directions  $\theta_{\omega_1} = \tilde{\theta}_1 = 32^\circ$  and  $\theta_{\omega_2}$  which changes on the different figures. On (a),  $\theta_{\omega_2} = 47^\circ$ , which corresponds to  $1/2(\tilde{\theta}_1 + \tilde{\theta}_2)$ ; on (d),  $\theta_{\omega_2} = \tilde{\theta}_2 = 62^\circ$ . In between, we set:  $\theta_{\omega_2} = 52^\circ$  and  $57^\circ$ . The values of the two components  $\gamma_{\omega_1}$  and  $\gamma_{\omega_2}$ , which are the only components allowed to be non-zero, are varying along the two axis.  $\lambda$  is fixed to 9.5, which belongs to the interval  $I$ . In each figure, we see four (local) minima: the local minimum at  $\mathbf{0}$ , local minima along the axis (i.e., one non-zero component), and those corresponding to strictly two non-zero components. The global minimum (black filled circle) is located on Figure 1d for  $\gamma_{\omega_1} = \gamma_{\omega_2} = 7$  and its value is  $2\lambda = 19$ .

### 3.3. Minimizers of $J_{\text{CEL0}}$

The global minimizer of  $J_{\ell_0}$  is preserved in  $J_{\text{CEL0}}$ , but the number of local minima of  $J_{\text{CEL0}}$  may be inferior to the number of local minima of  $J_{\ell_0}$ . Particularly, a local minimum  $\gamma$  of  $J_{\text{CEL0}}$  verifies  $|\gamma_i| \in \{0\} \cup [\sqrt{2\lambda}, +\infty)$ ; hence local minimizers of  $J_{\ell_0}$  having at least one component  $|\gamma_i| \in (0, \sqrt{2\lambda})$  are not local minimizers of  $J_{\text{CEL0}}$  [21]. Figure 1e–h represents the loss surfaces of  $J_{\text{CEL0}}$ . We observe the suppression of local minima on  $J_{\text{CEL0}}$  for  $\theta_{\omega_2}$  close to  $1/2(\tilde{\theta}_1 + \tilde{\theta}_2) = 47^\circ$ , for which  $0 < \gamma_{\omega_2} < \sqrt{2\lambda} = 4.36$ . Some local minima of  $J_{\ell_0}$  are also only critical points of  $J_{\text{CEL0}}$ . Moreover, the local minima that  $J_{\ell_0}$  has at  $\mathbf{0}$  is no longer one in  $J_{\text{CEL0}}$ , which is particularly interesting for the initialization of iterative optimization algorithms.

However, those good properties also have a disadvantage: it leads to “flat” minima, i.e., large connected regions where the error remains approximately constant. This is illustrated on Figure 2, comparing the minimum of  $J_{\ell_0}$  and  $J_{\text{CEL0}}$  as a function of  $\theta_{\omega_1}$  and  $\theta_{\omega_2}$  for two (or one) non-zero components. Numerous points of the CEL0 surface are approximately at the level of the local minima corresponding to  $\theta_{\omega_1} = \theta_{\omega_2} = 1/2(\tilde{\theta}_1 + \tilde{\theta}_2)$ . This behavior appears for close sources when this point is also close to the global minimum.



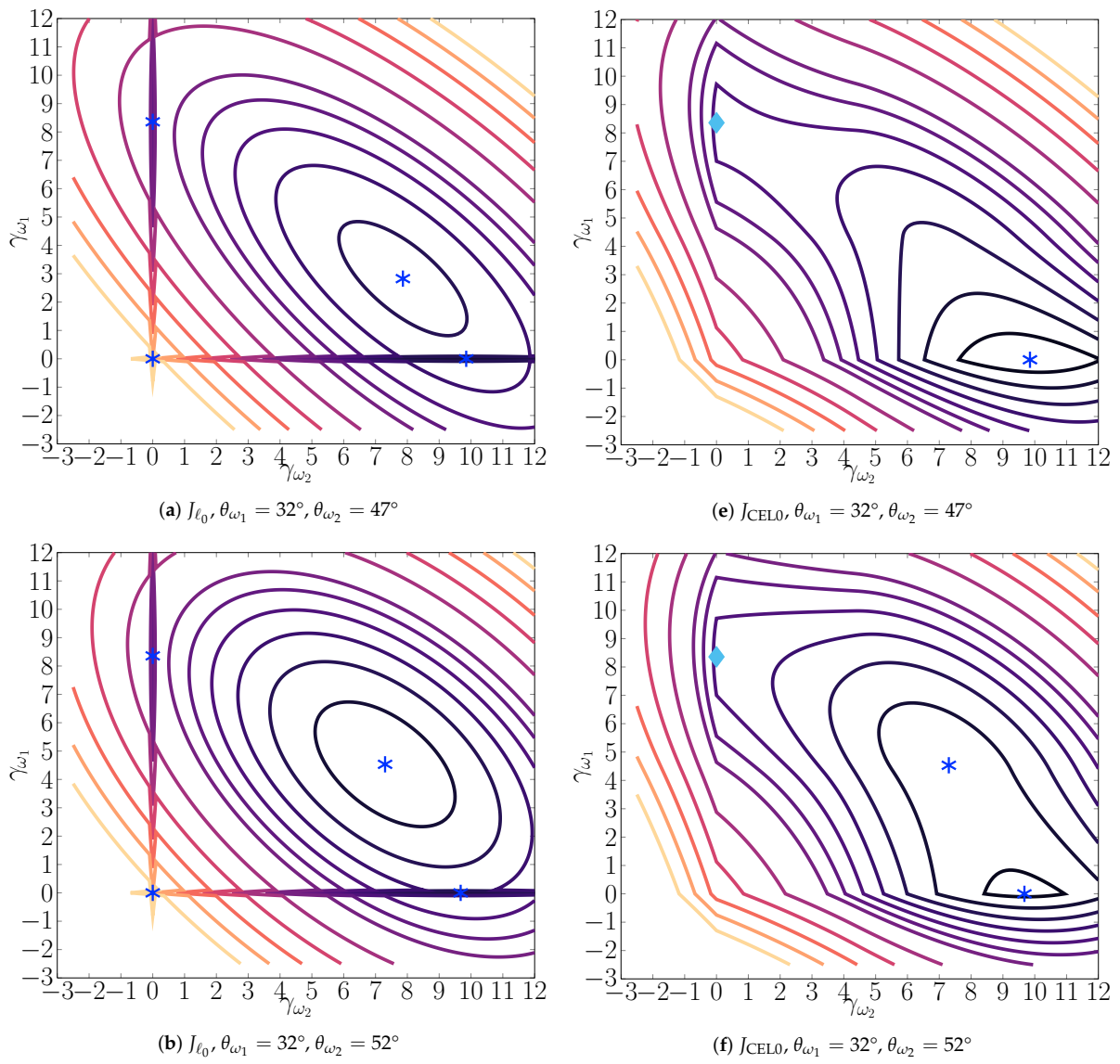
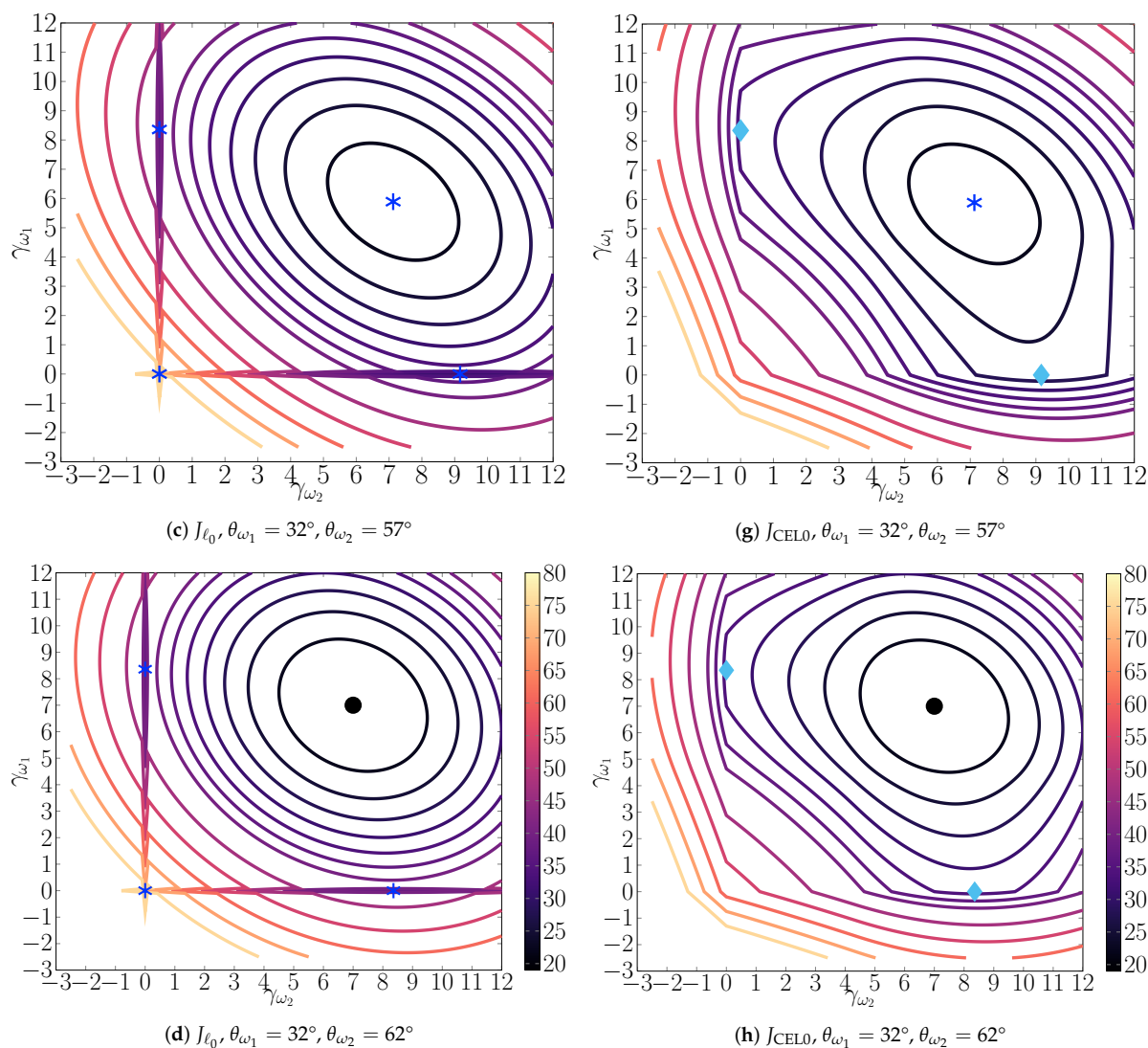
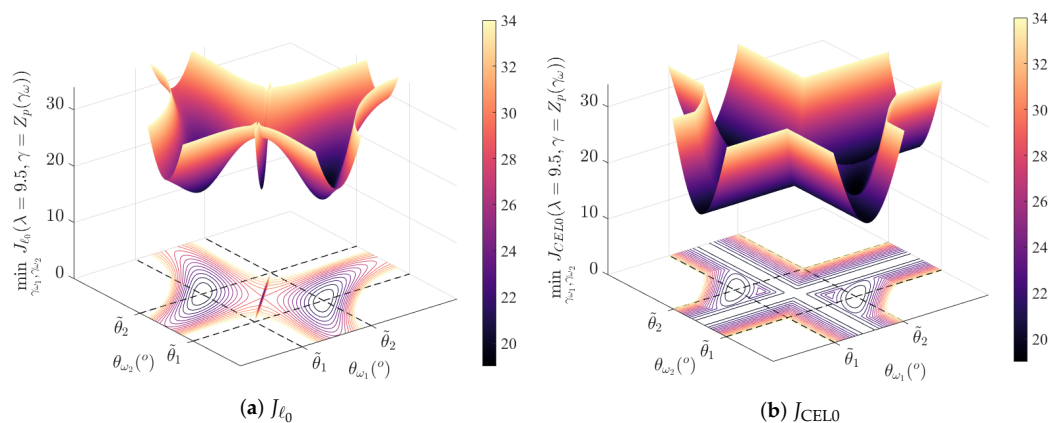


Figure 1. Cont.



**Figure 1.** Loss surfaces of  $J_{l_0}$  (a–d) and  $J_{CELO}$  (e–h) as a function of  $\gamma_{\omega_1}$  and  $\gamma_{\omega_2}$ , for  $\gamma = Z_p(\gamma_\omega)$ , with  $\#\omega = 2$ , i.e., at most two non-zero components corresponding to directions  $\theta_{\omega_1} = \tilde{\theta}_1$  and a varying  $\theta_{\omega_2}$ . When those directions correspond to the true ones  $\tilde{\theta}_1 = 32^\circ$  and  $\tilde{\theta}_2 = 62^\circ$  (d and h) the global minimum indicated by a black filled circle ( $\gamma_{\omega_1} = \gamma_{\omega_2} = 7$ ) is equal to  $2\lambda = 19$ . Local minima are indicated by the blue asterisks, while light blue diamonds represent critical points that are not local minima.



**Figure 2.** Minimum of the loss surfaces of  $J_{l_0}$  (a) and  $J_{CELO}$  (b) for  $\gamma = Z_p(\gamma_\omega)$ , with  $\#\omega = 2$ , as a function of  $\theta_{\omega_1}$  and  $\theta_{\omega_2}$ . The diagonal corresponds to  $\theta_{\omega_1} = \theta_{\omega_2}$ , i.e.,  $\#\omega = 1$ .



#### 4. Alternating between Loss Surfaces

For well separated sources, we have shown in [12] that IRL1 algorithm used to minimize  $J_{\text{CEL0}}$  (IRL1-CEL0) gives better statistical results than IHT and at a lower computational cost. Indeed, IRL1-CEL0 benefits from the suppression of local minima in this case. Unfortunately, IRL1-CEL0 fails for close sources. This behavior is illustrated on Figure 3a, for true sources at  $32^\circ$  and  $48^\circ$ . Let us note that MUSIC fails for such close sources. The IRL1-CEL0 algorithm is rapidly attracted by a local bad basin corresponding to a few non-zero components for directions in the middle of the true directions (lack of resolution). In this example, denoting  $\hat{\gamma}$  the final estimated vector and  $\hat{\omega}$  the set indicating the non-zero components, we verify that  $\mathbf{B}_{\hat{\omega}}^H \mathbf{B}_{\hat{\omega}} \hat{\gamma}_{\hat{\omega}} = \mathbf{B}_{\hat{\omega}}^H \mathbf{y}$  and  $\text{rank}(\mathbf{B}_{\hat{\omega}}) = \#\hat{\omega} = 5$ : it is a strict local minimum of  $J_{\ell_0}$ . The IHT algorithm also remains stuck in a local minimum with this time numerous non-zero components (Figure 3b), forming two clusters around the true directions. In order to avoid being attracted by a bad basin and take advantage of both regularizations, we propose to alternate the minimization between them. Based on this heuristic, we propose the optimization scheme ALICE-L0 (Alternated Landscapes Iterations for Complementary Enhancement for  $\ell_0$ ) detailed on Algorithm 1.

---

#### Algorithm 1. Optimization Scheme ALICE-L0 (Alternated Landscapes Iterations for Complementary Enhancement for $\ell_0$ )

---

**Input:** dictionary  $\mathbf{B}$ , observation  $\mathbf{y}$ ,  $\beta = 1/L$ ,  $L$  Lipschitz-constant of  $\mathbf{B}$ ,  $\tau_1$ ,  $\tau_2$ , stopping criteria

**Initialization:**  $\hat{\gamma}^{(0)} = \mathbf{0}$ ,  $i = 0$ ,  $i_{\text{outer}} = 0$

- $\epsilon^{(i)} = \|\mathbf{B}\hat{\gamma}^{(i)} - \mathbf{y}\|_2^2$

**while**  $\frac{|\epsilon^{(i-1)} - \epsilon^{(i-2)}|}{\epsilon^{(i-1)}} > \epsilon_{\text{lim}}$  and  $i_{\text{outer}} < n_{\text{lim}}$  **do**

- $i_{\text{outer}} = i_{\text{outer}} + 1$

- Compute  $\mathbf{w}$  weighting vector by:

$$w_g = (\sqrt{2\lambda} - |\gamma_g^{(i)}|) \mathbb{1}_{\{|\gamma_g^{(i)}| \leq \sqrt{2\lambda}\}}, g = 1..G$$

- $j = 1$ ,  $T^{(1)} = 1$ ,  $\mathbf{z}^{(1)} = \hat{\gamma}^{(i)}$

**while**  $\frac{|\epsilon^{(i-1)} - \epsilon^{(i-2)}|}{\epsilon^{(i-1)}} > \epsilon_{\text{lim},1}$  and  $j < n_{\text{lim},1}$  **do**

(weighted FISTA iterations)

- $j = j + 1$ ,  $i = i + 1$

- $\hat{\gamma}^{(i)} = \text{prox}_{\|\cdot\|_1, \lambda \beta \mathbf{w}} \left( \mathbf{z}^{(j-1)} - \beta \mathbf{B}^H (\mathbf{B} \mathbf{z}^{(j-1)} - \mathbf{y}) \right)$

- $T^{(j)} = \frac{1 + \sqrt{1 + (2T^{(j-1)})^2}}{2}$

- $\mathbf{z}^{(j)} = \hat{\gamma}^{(i)} + \frac{T^{(j-1)} - 1}{T^{(j)}} \left( \hat{\gamma}^{(i)} - \hat{\gamma}^{(i-1)} \right)$

**end while**

- $k = 0$

**while**  $\frac{|\epsilon^{(i-1)} - \epsilon^{(i-2)}|}{\epsilon^{(i-1)}} > \epsilon_{\text{lim},2}$  and  $k < n_{\text{lim},2}$  **do**

(IHT iterations)

- $k = k + 1$ ,  $i = i + 1$

- $\hat{\gamma}^{(i)} = \text{prox}_{\|\cdot\|_0, \lambda \beta} \left( \hat{\gamma}^{(i-1)} - \beta \mathbf{B}^H (\mathbf{B} \hat{\gamma}^{(i-1)} - \mathbf{y}) \right)$

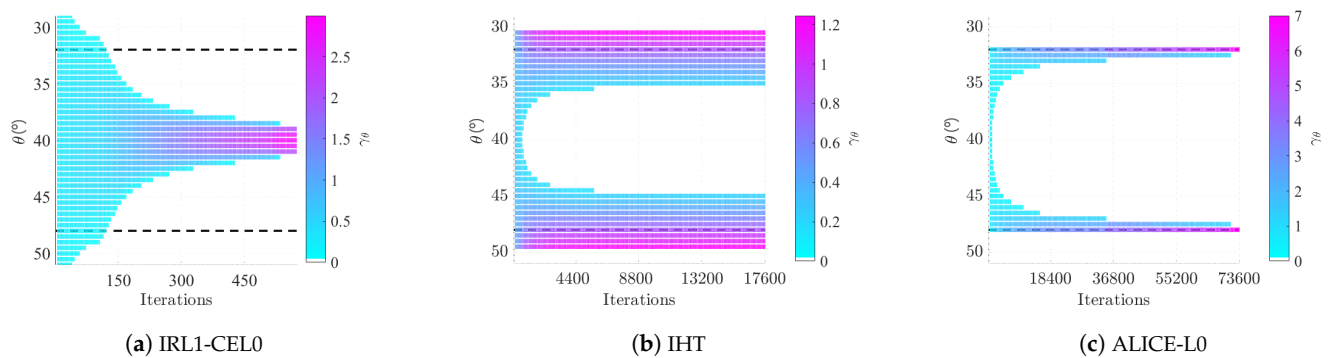
**end while**

- $\epsilon_{\text{lim},1} = \tau_1 \epsilon_{\text{lim},1}$ ,  $\epsilon_{\text{lim},2} = \tau_2 \epsilon_{\text{lim},2}$

**end while**

**return**  $\hat{\gamma}^{(i_{\text{end}})}$

---



**Figure 3.** Solutions as iterations go by for close sources with no noise, for  $\lambda = 0.78$ . X-axis: iteration number. Y-axis: directions associated with components of  $\hat{\gamma}$ . The color represents the level of the components. True sources  $\hat{\theta}_1 = 32^\circ$  and  $\hat{\theta}_2 = 48^\circ$  are indicated by the black dotted lines. Corresponding components of the optimal solution are equal to 7, all others are null.

We start the minimization considering the CEL0-regularized functional, and using  $\hat{\gamma}^{(0)} = \mathbf{0}$  as initialization. Indeed, we previously saw that this local minimum in  $J_{\ell_0}$  is suppressed in  $J_{\text{CEL0}}$ . Iterations of the weighted Fast Iterative Soft Thresholding Algorithm (weighted FISTA) aim to minimize the convex majorizer of the nonconvex CEL0 functional. For a weighting vector  $\mathbf{w}$ , iterations use the proximal of the weighted  $\ell_1$  function, defined by component by component as:

$$\left[ \text{prox}_{\|\cdot\|_1, \lambda\beta\mathbf{w}} \right]_g(\mathbf{x}) = \max \left\{ 0, 1 - \frac{\lambda\beta\mathbf{w}}{|x_g|} \right\} x_g \mathbb{1}_{\{|x_g| \neq 0\}}. \quad (11)$$

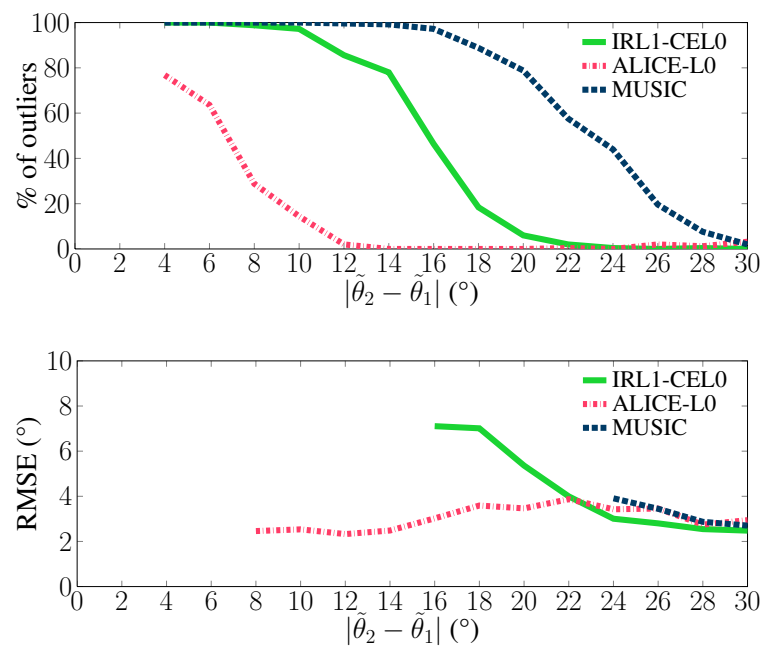
After some iterations, the estimated vector is used as initialization for IHT, which performs minimization steps over the  $\ell_0$ -regularized cost function. The hard threshold corresponds to the proximal of the  $\ell_0$ -norm, defined by:

$$\left[ \text{prox}_{\|\cdot\|_0, \lambda\beta} \right]_g(\mathbf{x}) = x_g \mathbb{1}_{\{|x_g| \geq \sqrt{\lambda\beta}\}}. \quad (12)$$

We then loop back to alternate between the loss surfaces. The behavior of our algorithm is represented on Figure 3c for close sources (in this article, we set  $\epsilon_{lim} = 1 \times 10^{-6}$ ,  $\epsilon_{lim,1} = 1 \times 10^{-2}$ ,  $\epsilon_{lim,2} = 1 \times 10^{-6}$ ,  $n_{lim} = 2000$ ,  $n_{lim,1} = 200$ ,  $n_{lim,2} = 200$ ,  $\tau_1 = 0.9$ ,  $\tau_2 = 1$ ). In this noiseless case, this is the only algorithm attaining the global minimum of  $J_{\ell_0}$ .

## 5. Statistical Performance

The purpose of this section is to numerically quantify the algorithms' performance as a function of the sources' separation. For that, two criteria will be used: the first one is the percentage of outliers (only one estimated direction or located at more than half a beamwidth of the true directions), and the second one the Root-Mean-Square-Error (RMSE) between estimated and true directions, calculated without outliers. The simulation setup is the one described in Section 3.1, for a Signal-to-Noise Ratio per source equal to 0 dB. Results are presented in Figure 4: we observe that the proposed scheme ALICE-L0 outperforms other methods in terms of statistical accuracy and resolution limit. Let us note that the behavior of IHT is unreliable with noise, and thus statistical results are not presented.



**Figure 4.** Performance as a function of the sources separation: percentage of outliers and RMSE (not reported when having more than 50% outliers). The regularization parameter  $\lambda$  is fixed to 0.78 according to the theoretical analysis [12].

## 6. Conclusions

We linked the operating limits of IHT and IRL1-CELO to the properties of corresponding loss landscapes in DOA estimation. To avoid the weaknesses of both criteria, an optimization scheme is proposed using alternatively  $J_{\ell_0}$  and  $J_{\text{CELO}}$ , i.e., alternating between the two regularizations. A particular implementation using  $\lambda$  obtained by [12,13] has been successfully tested, improving for example the resolution limit. Ongoing work concerns algorithm parameters (when to change of regularization) which are here left to the user.

**Author Contributions:** Conceptualization, A.D., A.F. and P.L.; methodology, A.D., A.F. and P.L.; software, A.D.; validation, A.D., A.F. and P.L.; formal analysis, A.D. and A.F.; investigation, A.D.; writing—original draft preparation, A.D.; writing—review and editing, A.D., A.F. and P.L.; visualization, A.D.; supervision, A.F. and P.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Agence Nationale de la Recherche et de la Technologie (2018/0698). The work reported in this manuscript has resulted in the French patent number 2107767.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Van Trees, H.L. *Optimum Array Processing; Part IV of Detection, Estimation, and Modulation Theory*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2002.
2. Avitabile, G.; Florio, A.; Coviello, G. Angle of Arrival Estimation Through a Full-Hardware Approach for Adaptive Beamforming. *IEEE Trans. Circuits Syst. II: Express Briefs* **2020**, *67*, 3033–3037. [[CrossRef](#)]
3. Stoica, P.; Nehorai, A. MUSIC, Maximum Likelihood, and Cramer-Rao Bound. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 720–741. [[CrossRef](#)]
4. Wang, W.; Wu, R. High Resolution Direction of Arrival (DOA) Estimation Based on Improved Orthogonal Matching Pursuit (OMP) Algorithm by Iterative Local Searching. *Sensors* **2013**, *13*, 11167–11183. [[CrossRef](#)] [[PubMed](#)]

5. Liu, J.; Zhou, W.; Juwono, F.H. Joint Smoothed l0-Norm DOA Estimation Algorithm for Multiple Measurement Vectors in MIMO Radar. *Sensors* **2017**, *17*, 1068. [[CrossRef](#)] [[PubMed](#)]
6. Yang, Z.; Li, J.; Stoica, P.; Xie, L. Chapter 11 - Sparse methods for direction-of-arrival estimation. In *Academic Press Library in Signal Processing*; Chellappa, R., Theodoridis, S., Eds.; Academic Press: Cambridge, MA, USA, 2018; Volume 7, pp. 509–581. [[CrossRef](#)]
7. Wu, X.; Zhu, W.; Yan, J. A high-resolution DOA estimation method with a family of nonconvex penalties. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4925–4938. [[CrossRef](#)]
8. Zhang, Z.; Wu, X.; Li, C.; Zhu, W.P. An  $\ell_p$ -Norm Based Method for Off-Grid DOA Estimation. *Circuits Syst. Signal Process.* **2019**, *38*, 904–917. [[CrossRef](#)]
9. Fan, Y.; Wang, J.; Du, R.; Lv, G. Sparse Method for Direction of Arrival Estimation Using Denoised Fourth-Order Cumulants Vector. *Sensors* **2018**, *18*, 1815. [[CrossRef](#)] [[PubMed](#)]
10. Soubies, E.; Chinatto, A.; Larzabal, P.; Romano, J.M.T.; Blanc-Féraud, L. Direction-of-Arrival Estimation Through Exact Continuous  $\ell_{2,0}$ -Norm Relaxation. *IEEE Signal Process. Lett.* **2021**, *28*, 16–20. [[CrossRef](#)]
11. Nikolova, M. Relationship between the optimal solutions of least squares regularized with L0-norm and constrained by k-sparsity. *Appl. Comput. Harmon. Anal.* **2016**, *41*, 237–265. [[CrossRef](#)]
12. Delmer, A.; Ferréol, A.; Larzabal, P. L0-Sparse DOA Estimation of Close Sources with Modeling Errors. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–21 January 2020.
13. Delmer, A.; Ferréol, A.; Larzabal, P. On Regularization Parameter for L0-Sparse Covariance Fitting Based DOA Estimation. In Proceedings of the ICASSP 2020, IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, 4–8 May 2020.
14. Marmin, A.; Castella, M.; Pesquet, J.C. How to Globally Solve Non-convex Optimization Problems Involving an Approximate L0 Penalization. In Proceedings of the ICASSP 2019, IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 12–17 May 2019; pp. 5601–5605. [[CrossRef](#)]
15. Candes, E.J. The Restricted Isometry Property and Its Implications for Compressed Sensing. *Comptes Rendus Math.* **2008**, *346*, 589–592. [[CrossRef](#)]
16. Selesnick, I. Sparse Regularization via Convex Analysis. *IEEE Trans. Signal Process.* **2017**, *65*, 4481–4494. [[CrossRef](#)]
17. Wen, F.; Chu, L.; Liu, P.; Qiu, R.C. A Survey on Nonconvex Regularization-Based Sparse and Low-Rank Recovery in Signal Processing, Statistics, and Machine Learning. *IEEE Access.* **2018**, *6*, [[CrossRef](#)]
18. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [[CrossRef](#)]
19. Soubies, E.; Blanc-Féraud, L.; Aubert, G. A Continuous Exact L0 penalty (CELO) for least squares regularized problem. *SIAM J. Imaging Sci.* **2015**, *8*, 1102492. [[CrossRef](#)]
20. Nikolova, M. Description of the minimizers of least squares regularized with  $\ell_0$ -norm. Uniqueness of the global minimizer. *SIAM J. Imaging Sci.* **2013**, *6*, 904–937. [[CrossRef](#)]
21. Soubies, E.; Blanc-Féraud, L.; Aubert, G. New Insights on the Optimality Conditions of the l2-l0 Minimization Problem. *J. Math. Imaging Vis.* **2020**, *62*, 808–824. [[CrossRef](#)]