Off-Grid Multisnapshot Spectrum Sensing for Cognitive Radar

AUGUSTO AUBRY, Senior Member, IEEE
Università degli Studi di Napoli Federico II, Napoli, Italy

PRABHU BABU
Indian Institute of Technology Delhi, New Delhi, Delhi, India

ANTONIO DE MAIO, Fellow, IEEE
Università degli Studi di Napoli Federico II, Napoli, Italy

LUCA PALLOTTA, Senior Member, IEEE
University of Basilicata, Potenza, Italy

Spectrum sensing is a key aspect of next-generation cognitive radars that make use of the perception-action cycle to improve their performance while endowing cohabitation with other systems. Awareness of the electromagnetic (EM) environment surrounding the radar is demanded to adapt its behavior to the changing scene. 2-D spectrum sensing is usually carried-out on uniformly spaced grids, over which the angle of arrival (AOA) of diverse (unknown) sources is estimated along with their frequency occupancy. To mitigate the performance degradations of on-grid methods, this article proposes an off-grid 2-D profile recovery strategy where the atoms are no longer fixed according to a given pool of nominal AOAs, but some flexibility is allowed to infer off-grid angle displacements. Hence, the angle-frequency profile recovery process is formalized as a regularized maximum likelihood estimation capable of exploiting the inherent block-sparsity of the overall profile. The resulting challenging optimization problem is handled through a maximum block improvement (MBI)-based method, which provides an estimate of the three variable blocks involved in the process, viz., noise power, 2-D profile, and angular displacements. Furthermore, in order to enhance the reliability of determining the space-frequency occupancy map and accurately estimating the angle displacements, three refinement strategies for the 2-D spectrum profile are suggested, suitably leveraging Bayesian information criterion and false discovery rate paradigms. The proposed framework is then validated through numerical simulations in some realistic EM environments, also comparing the three proposed refinement strategies.

Received 23 September 2024; revised 24 January 2025; accepted 9 February 2025. Date of publication 17 February 2025; date of current version 11 August 2025.

DOI. No. 10.1109/TAES.2025.3542747

Refereeing of this contribution was handled by W. Yi. see attached

The work of Augusto Aubry and Antonio De Maio was supported by the European Union in the NextGenerationEU plan through the Italian program "Bando PRIN 2022," D.D. 104/2022 (PE7, project "CIRCE," code E53D23000410006).

Authors' addresses: Augusto Aubry and Antonio De Maio are with the Department of Electrical Engineering and Information Technology, Università degli Studi di Napoli Federico II, I-80125 Napoli, Italy Email (augusto.aubry@unina.it; ademaio@unina.it), Prabhu Babu is with the CARE, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India Email (prabhubabu@care.iitd.ac.in), Luca Pallotta is with the Department of Engineering, University of Basilicata, 85100 Potenza, Italy Email (luca.pallotta@unibas.it). *(Corresponding author: Antonio De Maio.)*

I. INTRODUCTION

Spectral cognition plays a crucial role in next-generation cognitive radars, which leverage the perception-action cycle [1]. Given the characteristics of overlapping frequency emitters, a key objective of cognitive radars is to employ waveforms that enhance radar performance while maintaining spectral coexistence [2], [3], [4], [5], [6]. This requires electromagnetic (EM) awareness of the operational scenario, achieved through dynamic estimation of the spectrum occupancy, which is necessary for flexible and efficient use/management of the frequency resources [7]. In this context, many spectrum sensing algorithms have been developed, mainly in the field of communication networks, to address spectrum scarcity in certain frequency bands and to increase the degree of utilization of certain regions of the spectrum whose occupancy changes in time and space [8], [9], [10]. From the perspective of cognitive radar systems, obtaining precise information about the location of interference in both the frequency and angular domains enables more efficient use of space-time system resources. This accurate situational awareness allows the radar to dynamically allocate and optimize its operational parameters, boosting its performance while guaranteeing cohabitation with overlaid infrastructures. Furthermore, in a hostile environment the gathered 2-D spectral information is extremely useful to counter interference sources, via an appropriate shaping of the radar space-frequency response.

There are two major categories of approaches for spectrum sensing applications [11], [12]. The methods in the first group (a.k.a. supervised processes) take advantage of the inherent characteristics of signals transmitted by licensed emitters. For instance, waveform-based sensing algorithms, which correlate the measurements with known transmitter patterns, belong to this class [9]. Alternatively, unsupervised techniques can be employed, relying on statistical properties of the collected measurements, such as the data sample covariance matrix [13] or cyclostationary parameters [14] (e.g., condition number, largest eigenvalue, and trace). More recently, the authors in [15] and [16] devised two-dimentional (2-D) space-frequency sensing techniques that enforce block sparsity during the recovery phase, hence extending the sparse learning via iterative minimization (SLIM) algorithm developed in [17] to the block-sparsity scenario. It is worth noting that the SLIM has found a large application also in many other contexts [18], [19]. In some cognitive radar applications, the number of emitters surrounding the receiving system is far less than the

total number of space-frequency bins. As a result, a sparse model can be utilized to describe the received signal, and a sparsity-based signal recovery framework can be used to estimate emitter parameters, such as frequency occupancy (i.e., bandwidth) and location. Indeed, using sparsity on noisy inverse problems paves the way to the estimation of the unknown parameters with a better reliability than traditional methods, as also demonstrated with tests on real-recorded data in [20].

The above presented methodologies typically rely on a grid of possible unknowns, where each bin accounts for an angle of arrival (AOA). However, in many applications, grid selection is a hard task that in general leads to estimation errors [21] induced by the resulting mismatched signal model. To overcome the aforementioned limitation, off-grid strategies can be employed. In these approaches, the algorithm operates on a fixed grid while simultaneously estimating the grid offset along with the recovered signal [22], [23], [24], [25], [26], [27]. In addition, interpolation schemes can be used to minimize estimation errors [28], [29], or grid adjustments can be made dynamically during the estimation process [30]. In such cases, the devised techniques, which are often primarily focused on 1-D scenarios, can realize some approximations of the atoms. Generally, the off-grid direction-of-arrival (DOA) sparse recovery methods [21], [31] improve grid-based approaches by representing direction of arrival (DOA)s as deviations from predefined grid points. This family of techniques relies on a fine enough initial grid and an appropriate perturbation model, allowing for a balance between accuracy and computational efficiency: these methods represent a viable solution for real-time applications with moderate to high resolution requirements. In contrast, gridless sparse recovery (e.g., [32], [33], [34], [35]), works directly in the unconstrained state parameters, removing the need for a predefined grid. While this class of algorithm may provide superior accuracy and resolution, it requires higher computational demands [for instance, atomic-norm-based gridless techniques require (in general) solving a semidefinite program] and exhibits a strong sensitivity to array manifold mismatches (always present in practical applications [36, Ch. 5]) than the off-grid counterpart. Finally, gridless methods are sensitive to modeling errors [37], [38]. Other interesting gridless DOA estimation methods can be found in [39] and [40], where the Newton method is exploited for sinusoid signatures detection over an oversampled discrete Fourier transform grid followed by a specific parameters refinement strategy to face with manifold mismatches. The framework presented in this article belongs to the family of off-grid algorithms, being more general than gridless methods, such as those based on atomic norm [32], [33]. In practice, gridless methods usually require specific structures in the array manifold, such as Vandermonde in the case of uniform linear array, which unfortunately might not hold true in practice, leading to model mismatches and performance degradations.

To counteract the performance limitations discussed earlier and to fill the gap in [16], which relies on an on-grid setup, this article introduces an off-grid 2-D profile recovery strategy. Unlike methods that depend on a fixed set of nominal AOAs, this approach offers flexibility by adjusting the atoms to accommodate off-grid angle displacements. Consequently, the recovery process of the angle-frequency profile is framed as an regularized maximum likelihood (RML) estimation, aimed at exploiting the inherent block-sparsity of the entire profile. The resulting optimization problem is tackled using an minorization-maximization (MM)-maximum block improvement (MBI) approach. This method iteratively refines the estimates of three main variables, viz., noise power, 2-D profile, and angular displacements. Specifically, by iteratively focusing on each block, the MBI improves the the accuracy of the angle-frequency profile recovery. This approach simplifies the optimization process by breaking it into more manageable components and making use of the problem's inherent structure. In addition, to improve the reliability of determining the space-frequency occupancy map and to accurately estimate angle displacements, three refinement strategies for the 2-D spectrum profile are proposed. In the former, the profile refinement is performed using the Bayesian information criterion to obtain a bespoke selection of the dominant atoms. In contrast, the latter employs a two-stage approach for selecting the dominant atoms. The first stage involves a coarse selection, which can be performed either through orthogonal matching pursuit [41] or by analyzing the per-row energy profile of the estimated signal. Moreover, in both cases, the second stage involves refining the selection by applying the false discovery rate paradigm [42], [43]. Interesting case studies, accounting for realistic EM environments, are provided to validate the proposed framework, likewise comparing the three suggested strategies for 2-D spectrum refinement.

The main technical contributions of this article can be streamlined as follows.

1) The space-frequency profile recovery problem is formulated as a solution to the RML estimation problem, involving a joint optimization of noise power, 2-D profile, and angular displacements.
2) A novel algorithm based on MM-MBI is designed to tackle the inference problem. It involves solving three subproblems, each corresponding to one of three variable blocks, viz., noise power, 2-D profile, and angular displacements. Hence, the MBI updates the block that most decreases the objective function in each iteration, while keeping the others fixed.
3) The convergence properties of the devised algorithm are rigorously proved, showing that any cluster point of the generated sequence of estimates satisfies the Karush–Kuhn–Tucker conditions of the optimization problem at hand.
4) Three refinement strategies are proposed to boost jointly the reliability of the space-frequency occupancy map recovery process and the accuracy of the angle displacements.

5) The effectiveness of the novel sensing procedures is thoroughly assessed, along with an analysis of their computational complexity.

The rest of this article is organized as follows. Section II formulates the 2-D spectrum sensing problem under consideration. Section III lays out the proposed solution when the source signals possibly come from off-grid directions, describing the entire 2-D profile reconstruction procedure. In Section IV, three refinement strategies for the 2-D spectrum profile estimate are proposed. Section V presents some results obtained on simulated environments representing scenarios of practical interest. Finally, Section VI concludes this article.

NOTATIONS

We use boldface for vectors $\boldsymbol{a}$ (lower case) and matrices $\boldsymbol{A}$ (upper case). The transpose and the conjugate transpose operators are denoted by the symbols $(\cdot)^T$ and $(\cdot)^\dagger$, respectively. **diag** $(\boldsymbol{a})$ is the diagonal matrix whose ith diagonal element is the ith entry of $\boldsymbol{a}$, whereas $\mathrm{tr}(\boldsymbol{A})$ is the trace of the matrix $\boldsymbol{A}$. $\boldsymbol{I}$ refers to the identity matrix, 0 is the matrix with null entries (their size is determined from the context), whereas $\emptyset$ denotes the empty set. The all-ones column vector of size N is denoted by $\mathbf{1}_N$. $\mathbb{R}^N$ and $\mathbb{C}^N$ are the sets of N-dimensional vectors of real and complex numbers, respectively, whereas $\mathbb{N}$ stands for natural numbers. $\boldsymbol{A}^{-1}$ is the inverse of a square matrix $\boldsymbol{A}$, whereas $\boldsymbol{A}^+ = (\boldsymbol{A}^\dagger\boldsymbol{A})^{-1}\boldsymbol{A}^\dagger$ is the Moore–Penrose pseudo-inverse of the singular matrix $\boldsymbol{A}$. The symbols $\otimes$ and $\odot$ indicate the Kronecker and the elementwise or Hadamard product, respectively. The letter j represents the imaginary unit (i.e., $j = \sqrt{-1}$). For any complex number x, $|x|$ indicates its modulus, and $\mathrm{Re}\{x\}$ indicates its real part. $\|\cdot\|$ denotes the Frobenius matrix norm. $\mathbb{E}[\cdot]$ is the statistical expectation and χ^2_k denotes a χ^2-distributed random variable with k degrees of freedom. Finally, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{M})$ means that $\boldsymbol{x}$ is a circularly symmetric Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{M}$.

II. SYSTEM AND SIGNAL MODEL

Without loss of generality, let us consider a radar system exploiting a uniform linear array (ULA) of M antenna elements,[1] that collects the signals transmitted by K radio frequency (RF) emitting sources that are located at specific unknown angular directions, say $\bar{\theta}_1, \ldots, \bar{\theta}_K$, each with its own (unknown) bandwidth. Let us also assume that there is only one source arriving from a specific angular direction, as illustrated in the pictorial representation of Fig. 1.

As shown in [15] and [16], the baseband discrete-time signal at the output of the receiving array (obtained by sampling the continuous-time signals, on the sensing bandwidth B, according to the associated Nyquist rate) for the

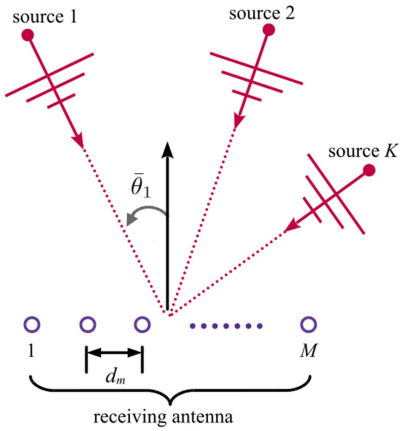


Fig. 1. Pictorial representation of the sensing scenario where a radar system, in passive mode, comprising multiple receiving antennas is used to acquire signals by multiple sources.

hth snapshot, $h = 1, \ldots, N_1$, with N_1 the total number of available snapshots, can be written as

$$\boldsymbol{y}_h = \sum_{k=1}^{K}\sum_{m=1}^{N_F} \bar{\boldsymbol{s}}(\bar{\theta}_k, \omega_m)\bar{a}_{k,m,h} + \boldsymbol{w}_h \tag{1}$$

where $N_F \geq N$ is the number of frequency bins, with N the number of available temporal observations per snapshot, $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N_1} \in \mathbb{C}^{NM}$ are the interference vectors affecting the different data-windows, modeled as independent and identically distributed circularly symmetric white Gaussian vectors, with zero mean and mean square value σ^2, and $\bar{a}_{k,m,h}$ is a scalar term proportional to the Fourier transform of the signal samples emitted by the kth source (during the hth data window) at the normalized frequency $\omega_m = (m-1)/N_F, m = 1, \ldots, N_F$. Moreover, $\bar{\boldsymbol{s}}(\theta, \omega)$ is the space-frequency steering vector associated with the angle θ and the frequency ω, i.e.,

$$\bar{\boldsymbol{s}}(\theta, \omega) = \boldsymbol{s}_F(\omega) \otimes \boldsymbol{s}(\theta) \tag{2}$$

with

$$\boldsymbol{s}_F(\omega) = \frac{1}{\sqrt{N}}\left[1, e^{j2\pi\omega}, \ldots, e^{j2\pi(N-1)\omega}\right]^T \in \mathbb{C}^N$$

and

$$\boldsymbol{s}(\theta) = \frac{1}{\sqrt{M}}\left[1, e^{j2\pi\frac{d_1}{\lambda_0}\sin(\theta)}, \ldots, e^{j2\pi\frac{d_{M-1}}{\lambda_0}\sin(\theta)}\right]^T \in \mathbb{C}^M$$

where $d_{\bar{m}} = \bar{m} \cdot d$, $\bar{m} = 0, \ldots, M-1$, d is the antenna interelement space, and λ_0 is the central carrier operating wavelength.

Before proceeding, let us introduce a grid of K_1 angular locations θ_i, $i = 1, \ldots, K_1$, chosen sufficiently dense such that there will be at most one source within each bin $[\theta_i - \delta/2, \theta_i + \delta/2]$. For the selection of the discretization step δ, for all the simulations conducted in this article, it is set as a fraction of the 3-dB beamwidth of the liner array pointing at boresight, namely,

$$\delta \leq \mathrm{asin}\left(\frac{0.891}{M}\right) \tag{3}$$

[1] The methodology proposed in this article can be utilized also for different array structures, as long as the expression of the array manifold is given, namely, the sensors position is known at the design stage.

where asin$(\cdot)$ denotes the arcsin function. To gather environmental awareness about space-frequency occupancy map, on-grid approaches suppose that the AOA of each source is a point of the assumed grid. Unfortunately, grid selection is a challenging problem and will inevitably result in grid mismatch [21]. In order to resolve the grid mismatch problem, off-grid sparse methods can be used instead. In these methods, a grid is still required for sparse estimation; however, unlike the on-grid methods, the AOA estimates are not restricted to the grid [21]. In such a case, the actual source AOA is modeled as the closest grid point plus a displacement, $\Delta\theta_i \in [-\delta/2, \delta/2]$, $i = 1, \ldots, K_1$. Otherwise stated, each AOA can be expressed as $\bar{\theta}_i = \theta_i + \Delta\theta_i$, $i = 1, \ldots, K_1$.

According to the above considerations, signal model (1), can be expressed in the following form

$$\boldsymbol{y}_h = \sum_{i=1}^{K_1} \sum_{m=1}^{N_F} \bar{\boldsymbol{s}}(\theta_i + \Delta\theta_i, \omega_m) a_{i,m,h} + \boldsymbol{w}_h \tag{4}$$

where $\bar{\boldsymbol{s}}(\theta_i + \Delta\theta_i, \omega_m)$, $i = 1, \ldots, K_1$, $m = 1, \ldots, N_F$, define the overall dictionary (to be learned according to the displacements' selection) and $a_{i,m,h}$, $i = 1, \ldots, K_1$, $m = 1, \ldots, N_F$, represent the unknown space-frequency profile at the hth snapshot. After establishing the values of $\Delta\theta_i$, $i = 1, \ldots, K_1$, a corresponding dictionary is derived in accordance with the structure of the signal model in (4). By refining the estimates of the variables $\Delta\theta_i \in [-\delta/2, \delta/2]$, a dynamic dictionary that adapts to the current environmental scenario is constructed.

To proceed further, let us assume δ is small enough such that the displacement effect on the actual steering vector in each bin can be modeled resorting to the first-order Taylor expansion of $\bar{\boldsymbol{s}}(\theta_i + \Delta\theta_i, \omega_m)$ around θ_i, i.e., [21], [24]

$$\bar{\boldsymbol{s}}(\theta_i + \Delta\theta_i, \omega_m) = \bar{\boldsymbol{s}}(\theta_i, \omega_m) + \dot{\boldsymbol{s}}(\theta_i, \omega_m)\Delta\theta_i, \;\; i = 1, \ldots, K_1 \tag{5}$$

with $\dot{\boldsymbol{s}}(\theta_i, \cdot)$ the derivative of $\bar{\boldsymbol{s}}(\theta_i, \cdot)$ with respect to the variable θ evaluated at $\theta = \theta_i$, whose expression is [44]

$$\dot{\boldsymbol{s}}(\theta_i, \omega_m) = \frac{\partial \bar{\boldsymbol{s}}(\theta_i, \omega_m)}{\partial \theta_i} = j\frac{2\pi}{\lambda_0}\cos(\theta_i) \cdot \boldsymbol{s}_F(\omega_m) \otimes \left(\boldsymbol{s}(\theta_i) \odot \left[d_0, \ldots, d_{M-1}\right]^T\right). \tag{6}$$

According to (5), the signal model in (4) boils down to

$$\boldsymbol{y}_h \approx \sum_{i=1}^{K_1} \sum_{m=1}^{N_F} \left(\bar{\boldsymbol{s}}(\theta_i, \omega_m) + \dot{\boldsymbol{s}}(\theta_i, \omega_m)\Delta\theta_i\right) a_{i,m,h} + \boldsymbol{w}_h \tag{7}$$

Note that, in general, the space-frequency profile is dependent on the snapshot index h, due to random variations of both source signals and channel state conditions (e.g., fading effects) among the different data windows. However, angular location and spectral support of the emitters are supposed stationary along the overall data acquisition time. Otherwise stated, if a space-frequency bin $(i^\star, m^\star)$ is inactive for the snapshot $h^\star$ (namely, $a_{i^\star,m^\star,h^\star} = 0$), then it is assumed that $a_{i^\star,m^\star,h} = 0$ for all recorded snapshots, that is

$$a_{i^\star,m^\star,h^\star} = 0 \quad \Rightarrow \quad a_{i^\star,m^\star,h} = 0 \quad \forall h = 1, \ldots, N_1. \tag{8}$$

The goal of 2-D spectrum sensing is to retrieve the overall space-frequency profile through the available measurements described via the signal model in (7). In particular, as in [16], the space-frequency occupancy map can be recovered from the estimated profile determining the angle-frequency bins whose total energy is nonzero, that is

$$\sum_{h=1}^{N_1} |a_{i^\star,m^\star,h}|^2 \neq 0. \tag{9}$$

Now, to better exploit the block-sparsity structure in the signal model (7), an equivalent (but more compact) matrix form of (7) is introduced. To this regard, let us recast (7) as

$$\boldsymbol{y}_h = \boldsymbol{H}(\boldsymbol{\Delta\theta})\boldsymbol{x}_h + \boldsymbol{w}_h, \;\; h = 1, \ldots, N_1 \tag{10}$$

where $\boldsymbol{x}_h \in \mathbb{C}^{K_1 N_F}$ is the vector containing the space-frequency profile for the hth snapshot, namely,

$$\boldsymbol{x}_h = \left[a_{1,1,h}, \ldots, a_{1,N_F,h}, a_{2,1,h}, \ldots, a_{K_1,N_F,h}\right]^T \tag{11}$$

whereas the matrix $\boldsymbol{H}(\cdot) \in \mathbb{C}^{NM \times K_1 N_F}$ corresponds to the model matrix (parameterized by the displacement angles $\Delta\theta_i$, $i = 1, \ldots, K_1$) whose structure is

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{\Delta\theta}) = \; & [\bar{\boldsymbol{s}}(\theta_1, \omega_1) + \dot{\boldsymbol{s}}(\theta_1, \omega_1)\Delta\theta_1, \ldots, \\ & \bar{\boldsymbol{s}}(\theta_1, \omega_{N_F}) + \dot{\boldsymbol{s}}(\theta_1, \omega_{N_F})\Delta\theta_1 \\ & \bar{\boldsymbol{s}}(\theta_2, \omega_1) + \dot{\boldsymbol{s}}(\theta_2, \omega_1)\Delta\theta_2, \ldots, \\ & \bar{\boldsymbol{s}}(\theta_{K_1}, \omega_{N_F}) + \dot{\boldsymbol{s}}(\theta_{K_1}, \omega_{N_F})\Delta\theta_{K_1}] \end{aligned} \tag{12}$$

with $\boldsymbol{\Delta\theta} = [\Delta\theta_1, \ldots, \Delta\theta_{K_1}]^T$.

REMARK 1 In (12), each *nominal* atom, as considered in the on-grid approach, is modified to account for the actual AOAs through the displacement term $\Delta\theta_i$.

When multiple snapshots are available, the overall signal collected at the receiver can be compactly written as

$$\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{\Delta\theta})\boldsymbol{X} + \boldsymbol{W} \tag{13}$$

with

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N_1}] \in \mathbb{C}^{MN \times N_1}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_{K_1 N_F} \end{bmatrix} = [\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{N_1}] \in \mathbb{C}^{K_1 N_F \times N_1}$$

and

$$\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N_1}] \in \mathbb{C}^{MN \times N_1}.$$

Equation (13) highlights that the received signal for each data-window is the sum of the weighted angular-frequency components pertaining to the different sources. This setup, usually referred to as the subspace signal model [45], [46], is widely exploited in signal processing applications (including AOA estimation and spectral analysis) where the functional form of the steering vectors is application-dependent.

Regardless of the actual steering vector structure, assuming $\boldsymbol{\Delta\theta}$ known, several techniques have been proposed in open literature to retrieve the unknown model parameters [47], [48], [49], [50].

III. PROBLEM FORMULATION AND PROPOSED ESTIMATION PROCEDURE

In this section, a space-frequency profile recovery framework leveraging on the MM-MBI paradigm is introduced. In particular, it can be obtained as the solution to an RML estimation problem. In this respect, a term fostering sparsity in the space-frequency map is added to the negative log likelihood function, associated with data model (13), of the unknown parameters σ^2, $\boldsymbol{X}$, and $\boldsymbol{\Delta\theta}$, and some constraints are enforced at the inference stage. Moreover, the same problem can also be obtained according to a Bayesian framework, leading to a maximum a posteriori estimate of the parameters σ^2, $\boldsymbol{X}$, and $\boldsymbol{\Delta\theta}$ under specific assumption on the prior, namely

$$\boldsymbol{y}_i|\boldsymbol{X}, \boldsymbol{\Delta\theta}, \sigma^2 \sim \mathcal{N}\left(\boldsymbol{H}(\boldsymbol{\Delta\theta})\boldsymbol{x}^i, \sigma^2\boldsymbol{I}\right), \quad i = 1, \ldots, N_1 \tag{14}$$

with $\boldsymbol{y}_i|\boldsymbol{X}, \boldsymbol{\Delta\theta}, \sigma^2$, $i = 1, \ldots, N_1$, statistically independent random vectors, $\sigma^2 \sim \mathcal{U}(\sigma_L^2, \sigma_U^2)$, $\Delta\theta_i \sim \mathcal{U}(-\delta/2, \delta/2)$, and

$$f_{\boldsymbol{X}}(\boldsymbol{X}) \propto \prod_{k=1}^{K_1 N_F} \exp\left\{-\frac{2}{q}\left[\left(\|\boldsymbol{x}_k\|^2 + \epsilon\right)^{q/2}\right]\right\}$$

with $\boldsymbol{x}_k$ the kth row of $\boldsymbol{X}$ and $\epsilon > 0$ a smoothing factor making (16) differentiable. Note that σ^2, $\boldsymbol{X}$, and $\boldsymbol{\Delta\theta}$ are statistically independent quantities, whereas $f_{\boldsymbol{X}}(\boldsymbol{X})$ is a block-sparsity promoting prior for $\boldsymbol{X}$. According to the above discussion and considerations, the environmental state estimation process can be formulated as follows:

$$\mathcal{P}\begin{cases} \min\limits_{\boldsymbol{X}, \sigma^2, \boldsymbol{\Delta\theta}} & NMN_1 \log(\sigma^2) \\ & \quad + \frac{1}{\sigma^2}\|\boldsymbol{H}(\boldsymbol{\Delta\theta})\boldsymbol{X} - \boldsymbol{Y}\|^2 + f_1(\boldsymbol{X}) \\ \text{s.t.} & \sigma_L^2 \le \sigma^2 \le \sigma_U^2 \\ & |\Delta\theta_i| \le \delta/2, \ i = 1, \ldots, K_1 \end{cases} \tag{15}$$

where

$$f_1(\boldsymbol{X}) = \sum_{k=1}^{K_1 N_F} \frac{2}{q}\left[\left(\|\boldsymbol{x}_k\|^2 + \epsilon\right)^{q/2} - 1\right] \tag{16}$$

is the block-sparsity promoting penalty term. The parameter q, $0 < q \le 1$, used in (16) is the quantity ruling the l_q-norm of the vector containing the space-frequency source energies (pushing for a sparse 2-D profile recovery). Specifically, as a general consideration, the lower q the larger the sparsity of the provided estimate [16], [17], [20]. In particular, when $q \to 0$, small values of $\|\boldsymbol{x}_k\|^2$ lead to arbitrarily low values of the considered penalty term. Otherwise stated, multiple null rows in the recovered matrix $\boldsymbol{X}$ are substantially forced. Notably, when $q = 1$, $\epsilon = 0$, and the angle displacements are fixed the recovery process demands the solution of the l_2/l_1-optimization program proposed in [51], [52], and [53] for noiseless measurements. Furthermore, if the noise power level is fixed, it is the group version of the basis pursuit denoising algorithm presented in [54] to perform the retrieval of block-sparse signals in the presence of noisy data. Moreover, in (15), σ_L^2 and σ_U^2 are, respectively, a lower bound and an upper bound for the interference background level. The former can be evaluated by characterizing the power level associated with the isolated operation of the receiver components [55], while the latter can be inferred through measurements under stressing conditions (e.g., related to the device operating temperatures) and taking into account a conservative confidence level estimate.

Problem $\mathcal{P}$ is a nonconvex and in general nopolinomial-hard optimization problem (the objective is indeed a nonconvex function). Nevertheless, a high quality solution to the formulated block-sparse recovery problem can be obtained resorting to the MM-MBI paradigm [56], [57], [58], where the noise variance σ^2, the space-frequency profile $\boldsymbol{X}$, and the off-grid angle displacements $\boldsymbol{\Delta\theta}$ are considered as variable blocks. In the MM-MBI framework, the optimization vector is first split into blocks of disjoint variables and the objective function (or its surrogate) is minimized with respect to each block while keeping unaltered the others. Hence, only the block ensuring the best result in terms of objective function is updated for the next iteration step. In other words, at each iteration, the MBI updates only the variable block providing the best objective improvement, while keeping the others settled to their previous values.

To proceed formally with reference to Problem $\mathcal{P}$, let us introduce the optimization variables,

$$\mathcal{Y} = \left\{\sigma^2, \boldsymbol{X}, \boldsymbol{\Delta\theta}\right\} \tag{17}$$

which are partitioned into three blocks, i.e., as already mentioned, σ^2, $\boldsymbol{X}$, and $\boldsymbol{\Delta\theta} = [\Delta\theta_1, \ldots, \Delta\theta_{K_1}]^T$. Hence, $\mathcal{Y}^{(n-1)}$ denotes the optimized variables at the $(n-1)$-th iteration, that is

$$\mathcal{Y}^{(n-1)} = \left\{\sigma^{2^{(n-1)}}, \boldsymbol{X}^{(n-1)}, \boldsymbol{\Delta\theta}^{(n-1)}\right\}. \tag{18}$$

Then, the MM-MBI procedure requires solving (possibly resorting to bespoke surrogate functions) the following three subproblems at each iteration to update the optimized solution [59], that is

$$\mathcal{P}_1\begin{cases} \min\limits_{\sigma^2} & NMN_1 \log(\sigma^2) \\ & \quad + \frac{1}{\sigma^2}\left\|\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)})\boldsymbol{X}^{(n-1)} - \boldsymbol{Y}\right\|^2 + f_1(\boldsymbol{X}^{(n-1)}) \\ \text{s.t.} & \sigma_L^2 \le \sigma^2 \le \sigma_U^2 \end{cases} \tag{19}$$

$$\mathcal{P}_2\begin{cases} \min\limits_{\boldsymbol{X}} & NMN_1 \log(\sigma^{2^{(n-1)}}) \\ & \quad + \frac{1}{\sigma^{2^{(n-1)}}}\left\|\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)})\boldsymbol{X} - \boldsymbol{Y}\right\|^2 + f_1(\boldsymbol{X}) \end{cases} \tag{20}$$

and

$$\mathcal{P}_3 \begin{cases} \min\limits_{\boldsymbol{\Delta\theta}} & NMN_1 \log(\sigma^{2^{(n-1)}}) \\ & + \frac{1}{\sigma^{2^{(n-1)}}} \left\| \boldsymbol{H}(\boldsymbol{\Delta\theta}) \boldsymbol{X}^{(n-1)} - \boldsymbol{Y} \right\|^2 + f_1(\boldsymbol{X}^{(n-1)}) \\ \text{s.t.} & |\Delta\theta_h| \leq \delta/2, \ h = 1, \ldots, K_1. \end{cases} \tag{21}$$

The solution to Problems $\mathcal{P}_1$ and $\mathcal{P}_2$ is discussed in Section III-A, whereas the solution to Problem $\mathcal{P}_3$ is derived in Section III-B.

A. Solution to Problems $\mathcal{P}_1$ and $\mathcal{P}_2$

Given $\boldsymbol{\Delta\theta}^{(n-1)}$, the solution to Problem $\mathcal{P}_1$ and $\mathcal{P}_2$ with respect to the variables σ^2 and $\boldsymbol{X}$, respectively, (where the other block is fixed at the previous instance), can be found with the procedure in [16], herein reported for completeness.

With reference to Problem $\mathcal{P}_1$, fixing $\boldsymbol{X}$ and $\boldsymbol{\Delta\theta}$, the optimal solution is obtained setting to zero the derivative and accounting for the constraint set, i.e.,

$$\sigma^{2^{(n)}} = \min\left(\max\left(\sigma_L^2, \hat{\sigma}^2\right), \sigma_U^2\right) \tag{22}$$

where

$$\hat{\sigma}^2 = \frac{1}{NMN_1} \|\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{X}^{(n-1)} - \boldsymbol{Y}\|^2. \tag{23}$$

Regarding Problem $\mathcal{P}_2$, σ^2 and $\boldsymbol{\Delta\theta}$ play the role of problem parameters. Hence, it can be handled resorting to the MM paradigm. Specifically, a tight upper bound to its objective function is

$$NMN_1 \log(\sigma^2) + \frac{\|\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{X} - \boldsymbol{Y}\|^2}{\sigma^{2^{(n-1)}}} + \|\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}} \boldsymbol{X}\|^2$$

where

$$\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}} = \mathbf{diag}\left(\frac{1}{\left(\|\boldsymbol{x}_1^{(n-1)}\|^2 + \epsilon\right)^{\frac{1}{2}-\frac{q}{4}}}, \ldots, \frac{1}{\left(\|\boldsymbol{x}_{K_1 N_F}^{(n-1)}\|^2 + \epsilon\right)^{\frac{1}{2}-\frac{q}{4}}} \right). \tag{24}$$

Hence, the update of the space-frequency profile is obtained solving

$$\mathcal{P}_2^1 \begin{cases} \min\limits_{\boldsymbol{X}} & \frac{\|\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{X} - \boldsymbol{Y}\|^2}{\sigma^{2^{(n-1)}}} + \|\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}} \boldsymbol{X}\|^2. \end{cases} \tag{25}$$

Note that, Problem $\mathcal{P}_2^1$ is a strictly convex optimization problem [60] whose optimal solution is

$$\begin{aligned} \boldsymbol{X}^{(n)} &= \left(\boldsymbol{H}^\dagger(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)}) + \sigma^{2^{(n-1)}} \boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^2\right)^{-1} \\ &\quad \boldsymbol{H}^\dagger(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{Y} \\ &= \boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2} \boldsymbol{H}^\dagger(\boldsymbol{\Delta\theta}^{(n-1)}) \\ &\quad \left(\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2} \boldsymbol{H}^\dagger(\boldsymbol{\Delta\theta}^{(n-1)}) + \sigma^{2^{(n-1)}} \boldsymbol{I}\right)^{-1} \boldsymbol{Y}. \end{aligned} \tag{26}$$

Notice that, since $\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)}) \boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2} \boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n-1)})^\dagger \succeq \boldsymbol{0}$ and $\sigma^{2^{(n-1)}} \boldsymbol{I} \succ \boldsymbol{0}$, the conjugate gradient least squares (LS) technique [61] could be used to efficiently compute (26).

B. Solution to Problem $\mathcal{P}_3$

In this subsection, a procedure to determine an optimal solution $\boldsymbol{\Delta\theta}$ to Problem $\mathcal{P}_3$, where $\boldsymbol{X}$ and σ^2 are fixed, is developed.

To this end, note that Problem $\mathcal{P}_3$ is a convex box constrained quadratic problem, and thus, it can be optimally handled leveraging on the MBI approach with respect to the set of variables $\{\Delta\theta_1, \ldots, \Delta\theta_{K_1}\}$. In this respect, let us denote by $\mathcal{Y}_1^{(n_1-1)}$ the optimized variables at the (n_1-1)th inner iteration,[2] that is

$$\mathcal{Y}_1^{(n_1-1)} = \left\{\Delta\theta_1^{(n_1-1)}, \Delta\theta_2^{(n_1-1)}, \ldots, \Delta\theta_{K_1}^{(n_1-1)}\right\}. \tag{27}$$

Then, the MBI approach requires solving the following K_1 subproblems at each iteration [59], namely,

$$\mathcal{P}_3^i \begin{cases} \min\limits_{\Delta\theta_i} & \left\| \boldsymbol{H}(\widetilde{\boldsymbol{\Delta\theta}}_i) \boldsymbol{X} - \boldsymbol{Y} \right\|^2 \\ \text{s.t.} & |\Delta\theta_i| \leq \delta/2 \end{cases} \tag{28}$$

for $i = 1, \ldots, K_1$, where

$$\begin{aligned} \widetilde{\boldsymbol{\Delta\theta}}_i = \Big[&\Delta\theta_1^{(n_1-1)}, \ldots, \Delta\theta_{i-1}^{(n_1-1)}, \Delta\theta_i, \Delta\theta_{i+1}^{(n_1-1)} \\ &\ldots, \Delta\theta_{K_1}^{(n_1-1)} \Big]^T \in \mathbb{R}^{K_1 \times 1}. \end{aligned} \tag{29}$$

The main steps involved in the MBI-based strategy to solve Problem $\mathcal{P}_3$ are summarized in Algorithm 1, where $g_1(\mathcal{Y}_1^{(n_1)})$ indicates the objective function in (28) evaluated at $\mathcal{Y}_1^{(n_1)} = \{\Delta\theta_1^{(n_1)}, \ldots, \Delta\theta_{K_1}^{(n_1)}\}$. Moreover, the procedure is executed until convergence (or a maximum number of iterations $n_{1_{\max}}$) is reached, i.e., $g_1(\mathcal{Y}_1^{(n_1-1)}) - g_1(\mathcal{Y}_1^{(n_1)}) < \xi_1$ or $n_1 = n_{1_{\max}}$.

According to [57], the objective function monotonically decreases and any cluster point (whose existence is ensured by the boundness of the constraint set) is a stationary point. Being the problem at hand convex, this implies that any cluster point is also an optimal solution.

As shown in Appendix A, the closed-form solution to Problem (28), for a given i, can be derived solving the following equivalent constrained minimization problem

$$\begin{cases} \min\limits_{\Delta\theta_i} & \|\boldsymbol{c}_i\|^2 (\Delta\theta_i)^2 + \|\boldsymbol{\zeta} - \boldsymbol{C}_{-i} \boldsymbol{\Delta\theta}_{-i}\|^2 \\ & -2\Delta\theta_i \mathrm{Re}\left[\boldsymbol{c}_i^\dagger \left(\boldsymbol{\zeta} - \boldsymbol{C}_{-i} \boldsymbol{\Delta\theta}_{-i}\right)\right] \\ \text{s.t.} & |\Delta\theta_i| \leq \delta/2 \end{cases} \tag{30}$$

whose global optimal solution is given by

$$\Delta\theta_i = \min\left(\frac{\delta}{2}, \max\left(-\frac{\delta}{2}, \frac{\mathrm{Re}\left[\boldsymbol{c}_i^\dagger \left(\boldsymbol{\zeta} - \boldsymbol{C}_{-i} \boldsymbol{\Delta\theta}_{-i}\right)\right]}{\|\boldsymbol{c}_i\|^2}\right)\right) \tag{31}$$

[2] As triggering solution, $\boldsymbol{\Delta\theta}^{(n-1)}$ is adopted.

Algorithm 1: MBI Based Strategy to Solve Problem $\mathcal{P}_3$.

Input. σ^2, $\boldsymbol{H}$, $\boldsymbol{X}$, ξ_1 and $n_{1_{\max}}$.
Initialization. Set $\boldsymbol{\Delta\theta} = \boldsymbol{0}$.
repeat
  $n_1 = n_1 + 1$.
  for $i = 1 : K_1$ **do**
    Estimate $\Delta\theta_i$ as in (31).
    Evaluate the optimal value to Problem $\mathcal{P}_3^i$, say $\bar{\nu}_i^{(n_1)}$.
  end for
  Select $i^\star = \arg \min_{i=1,\ldots,K_1} \bar{\nu}_i^{(n_1)}$.
  Update the solution

$$\mathcal{Y}_1^{(n_1)} = \left\{ \ldots, \Delta\theta_{i^\star-1}^{(n_1-1)}, \Delta\theta_{i^\star}^{(n_1)}, \Delta\theta_{i^\star+1}^{(n_1-1)}, \ldots \right\}.$$

until $g_1(\mathcal{Y}_1^{(n_1-1)}) - g_1(\mathcal{Y}_1^{(n_1)}) \geq \xi_1$ or $n_1 < n_{1_{\max}}$
Output. $\widehat{\boldsymbol{\Delta\theta}} = \boldsymbol{\Delta\theta}^{(n_1)}$.

with

1) $\boldsymbol{C}_{-i} \in \mathbb{C}^{MNN_1 \times (K_1-1)}$, the matrix obtained from $\boldsymbol{C} \in \mathbb{C}^{MNN_1 \times K_1}$ [whose expression is provided in (50) of Appendix A] removing its ith column, i.e., the vector $\boldsymbol{c}_i \in \mathbb{C}^{MNN_1 \times 1}$ corresponding to the variable $\Delta\theta_i$;
2) $\boldsymbol{\Delta\theta}_{-i} \in \mathbb{C}^{(K_1-1)\times 1}$ the vector containing all the entries of $\boldsymbol{\Delta\theta}$ except that in position i;
3) $\boldsymbol{\zeta}$ is provided in (48) of Appendix A.

REMARK 2 Problem $\mathcal{P}$ can be further modified by adding a penalty $\mu\|\boldsymbol{\Delta\theta}\|$ promoting the estimates of not-null terms in the vector $\boldsymbol{\Delta\theta}$. This penalty slightly alters the objective function in Problem $\mathcal{P}$, but for low μ values, it makes Problem $\mathcal{P}_3$ strictly convex.

C. MM-MBI Based Strategy for 2-D Spectrum Sensing

The MM-MBI-based procedure developed in this article for off-grid 2-D spectrum sensing is summarized in Algorithm 2.

Note that the procedure in Algorithm 2 is executed until convergence (or a maximum number of iterations $n_{\max}$) is reached, i.e., $g(\mathcal{Y}^{(n-1)}) - g(\mathcal{Y}^{(n)}) < \xi$ or $n = n_{\max}$, with $g(\mathcal{Y}^{(n)}) = NMN_1 \log(\sigma^{2^{(n)}}) + \frac{1}{\sigma^{2^{(n)}}} \|\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n)})\boldsymbol{X}^{(n)} - \boldsymbol{Y}\|^2 + f_1(\boldsymbol{X}^{(n)})$.

As to the convergence analysis, some relevant properties of Algorithm 2 are summarized in Proposition 1.

PROPOSITION 1 Algorithm 2 enjoys the following properties.

1) Problem (15) is solvable,[3] $g(\mathcal{Y}^{(n)}) \geq \nu(\mathcal{P})$, and the sequence of estimates $\mathcal{Y}^{(n)}$ generated by Algorithm 2 is bounded.

[3] A minimization problem is solvable if it is feasible, bounded below, and its optimal value $\nu(\mathcal{P})$ is attained [62].

Algorithm 2: MM-MBI Based Strategy for Off-Grid 2-D Spectrum Sensing.

Input. σ_L^2, σ_U^2, ϵ, $\boldsymbol{H}$, $\boldsymbol{Y}$, $\boldsymbol{X}^{(0)}$, ξ, $n_{\max}$, and $q \in [0, 1]$.
Initialization. Set $n = 0$, $\boldsymbol{\Delta\theta}^{(0)} = \boldsymbol{0}$, and $\sigma^{2^{(0)}} = \min(\max(\sigma_L^2, \hat{\sigma}^2), \sigma_U^2)$.
repeat
  $n = n + 1$.
  Compute $\sigma^{2^{(n)}} = \min(\max(\sigma_L^2, \hat{\sigma}^2), \sigma_U^2)$ with $\hat{\sigma}^2$ specified in (23), and evaluate the objective value $\nu_1^{(n)}$ of Problem $\mathcal{P}_1$.
  Set $\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}$ as in (24).
  Compute $\boldsymbol{X}^{(n)}$ applying (26) and evaluate the objective value $\nu_2^{(n)}$ of Problem $\mathcal{P}_2$.
  Compute an optimal solution to Problem $\mathcal{P}_3$, $\boldsymbol{\Delta\theta}^{(n)}$, with the inner MBI-based technique described in Algorithm 1, and evaluate the objective value $\nu_3^{(n)}$ of Problem $\mathcal{P}_3$.
  Select $k^\star = \arg \min_{k=1,\ldots,3} \nu_k^{(n)}$.
  if $k^\star = 1$ **then**
    Update the solution

$$\mathcal{Y}^{(n)} = \left\{ \sigma^{2^{(n)}}, \boldsymbol{X}^{(n-1)}, \boldsymbol{\Delta\theta}^{(n-1)} \right\}.$$

  else if $k^\star = 2$ **then**
    Update the solution

$$\mathcal{Y}^{(n)} = \left\{ \sigma^{2^{(n-1)}}, \boldsymbol{X}^{(n)}, \boldsymbol{\Delta\theta}^{(n-1)} \right\}.$$

  else
    Update the solution

$$\mathcal{Y}^{(n)} = \left\{ \sigma^{2^{(n-1)}}, \boldsymbol{X}^{(n-1)}, \boldsymbol{\Delta\theta}^{(n)} \right\}.$$

    and the model matrix $\boldsymbol{H}(\boldsymbol{\Delta\theta}^{(n)})$.
  end if
until $g(\mathcal{Y}^{(n-1)}) - g(\mathcal{Y}^{(n)}) \geq \xi$ or $n < n_{\max}$
Output. $\hat{\sigma}^2 = \sigma^{2^{(n)}}$, $\hat{\boldsymbol{X}} = \boldsymbol{X}^{(n)}$, and $\widehat{\boldsymbol{\Delta\theta}} = \boldsymbol{\Delta\theta}^{(n)}$.

2) The objective $g(\mathcal{Y}^{(n)})$ is monotonically decreasing and converges to a finite value $g^\star$. Moreover, for any cluster point $\mathcal{Y}^\star$ of $\mathcal{Y}^{(n)}$, $g(\mathcal{Y}^{(\star)}) = g^\star$.
3) Any cluster point of the produced sequence of estimates is a stationary solution to Problem $\mathcal{P}$, provided that at each iteration an optimal solution to Problem $\mathcal{P}_3$ is obtained, i.e., the involved inner loop, defined by Algorithm 1, reaches convergence to a Karush-Kuhn-Tucker (KKT) point.

PROOF See Appendix B. □

IV. 2-D SPECTRUM PROFILE ESTIMATE REFINEMENT STRATEGIES

Leveraging the output provided by Algorithm 2, in this section some strategies are developed to boost the capabilities of reliably determining the space-frequency occupancy

map as well as of accurately estimating the angle displacements. In a nutshell, valuable procedures performing a refinement of the 2-D spectrum profile estimated according to the RML framework are now proposed, by means of tailored screening approaches.

Each refinement technique operates according to the following steps.

1) The vector containing the displacement estimates at the output of Algorithm 2, $\widehat{\boldsymbol{\Delta\theta}}$, is used to construct the new updated dictionary. Precisely, for each bin, the corresponding atom is properly redrawn by correcting the on-grid angle with the estimated displacement according to ideal steering structure of (2). Hence, the complete model matrix $\mathring{\boldsymbol{H}}(\widehat{\boldsymbol{\Delta\theta}})$ represents the learned dictionary obtained from Algorithm 2.
2) Identification of the active atoms in $\mathring{\boldsymbol{H}}(\widehat{\boldsymbol{\Delta\theta}})$ via model selection rules aimed at providing an estimate $\mathcal{R}$ of the sources support, and definition of the resulting source subspace $\hat{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$.
3) Estimation of the 2-D space-frequency profile $\tilde{\boldsymbol{X}}$ resorting to the LS paradigm to infer the behavior of the of $\hat{\boldsymbol{X}}$ corresponding to the active atoms, i.e., the columns of $\hat{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$, while setting to zero the remaining rows.
4) Update of the angle displacement estimates executing Algorithm 1 with $\tilde{\boldsymbol{X}}$ and $\hat{\boldsymbol{H}}(\widehat{\boldsymbol{\Delta\theta}})$ as inputs.

In the next subsections, three different strategies are proposed to perform the selection of the active atoms (i.e., handle the second item of the above list of steps) with a reduced computational cost.

A. BIC-Based Refinement

The technique described in this section, inspired by the approach developed in [16], resorts to a Bayesian information criterion (BIC)-based strategy to pick up the most relevant atoms from the learned dictionary. The idea is to gradually add the atoms, accounting for their relevance. To this end, they are first ordered according to their energy (namely, the energy of the signal associated with the specific atom), and then a BIC based-strategy is employed to specify an estimate of the set of the selected active row indexes. More precisely, let $\hat{\boldsymbol{X}}$ be the profile recovered by Algorithm 2, $\hat{\boldsymbol{X}}_o$ be the matrix obtained from $\hat{\boldsymbol{X}}$ sorting its rows such that

$$\|\hat{\boldsymbol{x}}_{o,1}\|^2 \geq \|\hat{\boldsymbol{x}}_{o,2}\|^2 \geq \cdots \geq \|\hat{\boldsymbol{x}}_{o,K_1N_F}\|^2 \tag{32}$$

namely, the per-row energy of $\hat{\boldsymbol{X}}_o$ is arranged in decreasing order, and

$$\boldsymbol{r}_o = [r_o(1), r_o(2), \ldots, r_o(K_1N_F)]^T \in \mathbb{N}^{K_1N_F}$$

be the vector containing the corresponding ordered row indexes, i.e., $\hat{\boldsymbol{X}}_{o,i} = \hat{\boldsymbol{X}}_{r_o(i)}$, $i = 1, \ldots, K_1N_F$. Then,

$$\mathcal{R} = \left\{r_o(1), \ldots, r_o(k^\star)\right\} \tag{33}$$

where

$$k^\star = \arg \min_{k\in\{1,\ldots,\bar{K}\}} \mathrm{BIC}(k) \tag{34}$$

with

$$\mathrm{BIC}(k) = 2NMN_1 \log\left(\|\mathring{\boldsymbol{H}}_o^k \hat{\boldsymbol{X}}_o^k - \boldsymbol{Y}\|^2\right) + (2N_1 + 2)k \log(2NMN_1), \quad k = 1, \ldots, \bar{K}. \tag{35}$$

Note that, in (35)

1) $\mathring{\boldsymbol{H}}_o^k \in \mathbb{C}^{MN,k}$ is the matrix containing the first k columns of $\mathring{\boldsymbol{H}}_o$, with $\mathring{\boldsymbol{H}}_o$ the matrix obtained from $\mathring{\boldsymbol{H}}$ sorting its columns according to the permutation induced by the vector $\boldsymbol{r}_o$;
2) $\hat{\boldsymbol{X}}_o^k$ is the matrix containing the first k rows of $\hat{\boldsymbol{X}}_o$;
3) $\bar{K}$ is an upper limit on the actual number of space-frequency sources (which stems from some upper bounds on the number of sources K and their frequency support);
4) the second term of (35) represents the BIC-based penalty.

Finally, $\tilde{\boldsymbol{X}}$ is derived as the least squares (LS) estimate of $\boldsymbol{X}$ associated with the selected active rows $\mathcal{R}$ of the updated learned dictionary $\hat{\boldsymbol{H}}(\boldsymbol{\Delta\theta}) = \mathring{\boldsymbol{H}}_{r_o(i)}^{k^\star}$, $i = 1, \ldots, k^\star$.

In what follows, Algorithm 2 enjoying the BIC-based refinement is denoted by off-grid MM-MBI BIC-based 2-D spectrum sensing (shortly OG-MBI-BIC).

B. FDR-Based Refinements

To find the active atoms in $\mathring{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$, a two-stage procedure, comprising a coarse and a fine screening, is herein introduced. The former faces with the selection of at most C columns of $\mathring{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$ that likely span the useful signal subspace. Otherwise stated, a matrix $\breve{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$ is constructed by suitably picking up the dominant atoms (for the description of the received signal) from $\mathring{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$. The latter eventually establishes the actual active atoms by applying the false discovery rate (FDR) algorithm outlined in [43].

Precisely, starting from the dominant atoms collected in $\breve{\boldsymbol{H}}(\widehat{\boldsymbol{\Delta\theta}})$, the LS estimate of $\boldsymbol{X}$ is first computed as

$$\breve{\boldsymbol{X}} = \breve{\boldsymbol{H}}^{+}\boldsymbol{Y} \tag{36}$$

and its approximate error covariance matrix estimate is evaluated as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{NMN_1}\|\boldsymbol{Y} - \breve{\boldsymbol{H}}\breve{\boldsymbol{X}}\|^2 \left(\breve{\boldsymbol{H}}^{\dagger}\breve{\boldsymbol{H}}\right)^{-1}. \tag{37}$$

Therefore, the problem of estimating $\mathcal{R}$ can be viewed as testing the following κ null hypotheses:

$$\begin{cases} H_1 & : & \breve{\boldsymbol{x}}_1 = 0 \\ H_2 & : & \breve{\boldsymbol{x}}_2 = 0 \\ \vdots & : & \vdots \\ H_\kappa & : & \breve{\boldsymbol{x}}_\kappa = 0. \end{cases} \tag{38}$$

In this regard, the test statistics

$$T_k \triangleq \frac{\|\breve{\boldsymbol{x}}_k\|^2}{\hat{\boldsymbol{\Sigma}}(k,k)} \tag{39}$$

Algorithm 3: FDR for 2-D Spectrum Sensing.

1: **Input.** $\breve{\boldsymbol{H}}$ and $\boldsymbol{Y}$.
2: Compute the LS estimate of $\boldsymbol{X}$ as in (36).
3: Compute $\hat{\boldsymbol{\Sigma}}$ as in (37).
4: Compute the test statistics T_k, $k = 1, \ldots, \kappa$ using (39).
5: Sort T_k, $k = 1, \ldots, \kappa$, in decreasing order, viz., $T_{[1]} \geq \ldots \geq T_{[\kappa]}$.
6: Compute the lth probability P_l as

$$P_l = \frac{\alpha\, l}{\kappa \eta_\kappa}, \;\; l = 1, \ldots, \kappa$$

with $\eta_\kappa = \log(\kappa) + 0.577$ and α the FDR.
7: Compute the lth threshold from a $\chi^2_{N_1}$ distribution with N_1 degrees of freedom as

$$\omega_{P_l} : \text{Prob}\left[T_{[l]} \geq \omega_{P_l} | H_{[l]}\right] = P_l.$$

8: Find

$$\hat{l} \stackrel{\triangle}{=} \max\left[l : T_{[i]} \geq \omega_{P_i}, i = 1, \ldots, l\right].$$

9: Reject the null hypotheses $H_{[i]}$ for $i = 1, \ldots, \hat{l}$ and accept $H_{[\hat{l}+1]}, \ldots, H_{[\kappa]}$
10: Estimate the support as $\mathcal{R}_{\text{FDR}} = \{[1], [2], \ldots, [\hat{l}]\}$.
11: **Output.** $\hat{\boldsymbol{H}}(\boldsymbol{\Delta\theta}) = \breve{\boldsymbol{H}}_{\mathcal{R}_{\text{FDR}}}$.

are employed, where $\hat{\boldsymbol{\Sigma}}(k, k)$ is the (k, k)th entry of $\hat{\boldsymbol{\Sigma}}$. Under null hypotheses, T_k can be approximated with a $\chi^2_{N_1}$ random variable. Now, the FDR algorithm can be implemented as summarized in Algorithm 3. Finally, the dictionary matrix screened by the FDR $\hat{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$ is used to recover the 2-D profile containing only the effective atoms, i.e., $\tilde{\boldsymbol{X}}$.

Let us now focus on the coarse screening stage, which is accomplished by employing either the orthogonal matching pursuit (OMP) (with rows sparsity) [41], or the per-row energy profile of $\hat{\boldsymbol{X}}$.

1) *OMP:* The OMP procedure [41] is herein described with respect to its extension to the row-sparsity case. It is a regressor selection algorithm which preselects a number of regressors and obtains the LS estimates of the corresponding coefficients. Being it an iterative greedy selection process, at each iteration, it selects the most dominant regressors from the available set [41]. The main steps involved in the OMP technique are summarized in Algorithm 4, where κ denotes an upper bound to the number of dominant atoms, $\mathring{\boldsymbol{h}}_k$ is the kth column of the matrix $\mathring{\boldsymbol{H}}(\boldsymbol{\Delta\theta})$, and $\mathcal{R}^i = \{r(1), \ldots, r(i)\}$ indicates the space-frequency support estimate up to the ith iteration. Hence, the model matrix at the output of the OMP is given by $\breve{\boldsymbol{H}}(\boldsymbol{\Delta\theta}) = \mathring{\boldsymbol{H}}_{\mathcal{R}^\kappa}$.

Algorithm 2 along with the OMP-FDR-based refinement is indicated as off-grid MM-MBI OMP-FDR-based 2-D spectrum sensing (OG-MBI-OFDR).

2) *Energy Profile:* The second coarse identification strategy of the dominant atoms performs the selection of the local peaks of the per-row energy profile of the signal at the output of Algorithm 2, i.e., $\hat{\boldsymbol{X}}$. More precisely, for each space-frequency bin the mean power is evaluated as

Algorithm 4: OMP for 2-D Spectrum Sensing.

1: **Input.** $\mathring{\boldsymbol{H}}$, $\boldsymbol{Y}$, and κ.
2: **Initialization.** Set $\boldsymbol{R}_0 = \boldsymbol{Y}$, and $\mathcal{R}^0 = \emptyset$.
3: **for** $i = 1 : \kappa$. **do**
4: $\quad k_i = \arg \max_{k \in [1, \ldots, K_1 N_F] - \mathcal{R}^{i-1}} \|\mathring{\boldsymbol{h}}_k^\dagger \boldsymbol{R}_{i-1}\|^2$
5: $\quad \mathcal{R}^i = \mathcal{R}^{i-1} \cup k_i$
6: $\quad \boldsymbol{R}_i = (\boldsymbol{I} - \mathring{\boldsymbol{H}}_{\mathcal{R}^i} (\mathring{\boldsymbol{H}}_{\mathcal{R}^i}^\dagger \mathring{\boldsymbol{H}}_{\mathcal{R}^i})^{-1} \mathring{\boldsymbol{H}}_{\mathcal{R}^i}^\dagger) \boldsymbol{Y}$
7: **end for**
8: **Output.** $\breve{\boldsymbol{H}}(\boldsymbol{\Delta\theta}) = \mathring{\boldsymbol{H}}_{\mathcal{R}^\kappa} \in \mathbb{C}^{NM \times \kappa}$.

the energy in each row in the matrix $\hat{\boldsymbol{X}}$ associated with the related atom and normalized to the number of snapshots, that is

$$\hat{p}_i = \frac{\|\hat{\boldsymbol{x}}_i\|^2}{N_1}, \;\; i = 1, \ldots, K_1 N_F \tag{40}$$

where the index i is associated to a specific space-frequency bin. Then, the local peaks (i.e., a data point strictly greater than its immediate neighbors comprising eight surrounding angle-frequency bins) are extracted from the space-frequency mean power map, providing a new reduced size support, say $\mathcal{R}^E = \{r^E(1), \ldots, r^E(K_E)\}$, with $r^E(i)$ the ith location of a detected peak, $K_E \leq \overline{K}_E$ the number of detected peaks (with K_E not related in any way to the iterations of the OMP), and $\overline{K}_E$ an upperbound on it. Finally, the model matrix at the output of this screening is given by $\breve{\boldsymbol{H}}(\boldsymbol{\Delta\theta}) = \mathring{\boldsymbol{H}}_{\mathcal{R}^E} \in \mathbb{C}^{NM \times K_E}$.

Algorithm 2 exploiting the per-row energy profile-FDR-based refinements is referred as off-grid MM-MBI energy-FDR-based 2-D spectrum sensing (OG-MBI-EFDR).

C. Computational Complexity

The computational complexity[4] of the proposed refinement strategies is summarized in Table I. The computational cost of the BIC-based strategy is dominated by the evaluation of (35). Regarding the strategies based on the FDR, it is worth noting that Algorithm 3 does not provide a significant contribution to the final complexity. Hence, the only contribution comes from OMP (in the first FDR-based refinement strategy), which is in the steps 4 and 6 of Algorithm 4 and is iterated κ times. From the above considerations, it can be concluded that the overall complexity of the per-row energy profile-FDR-based approach can be assumed negligible.

V. PERFORMANCE ASSESSMENT

The purpose of this section is to assess the effectiveness of the proposed procedure in terms of 2-D spectrum sensing capabilities, i.e., the reliable recovery of the space-frequency occupancy map and off-grid displacement estimation. To this end, the three approaches OG-MBI-BIC, OG-MBI-FDR, and OG-MBI-EFDR described in

[4] The computational complexity is quantified making use of the usual Landau notation $\mathcal{O}(n)$. In this respect, an algorithm is $\mathcal{O}(n)$ if its implementation requires a number of flops proportional to n [63].

TABLE I
Computational Complexity of the Proposed 2-D Profile Refinement Strategies

algorithm	computational complexity
BIC	$\mathcal{O}\left(K_1 N_F N M N_1 \bar{K}\right)$
OMP-FDR	$\mathcal{O}\left(K_1 N_F N^2 M^2 N_1 \kappa\right)$
Energy profile-FDR	$\mathcal{O}\left(K_1 N_F \overline{K}_E\right)$

Section IV are considered. In addition, the on-grid BIC-based block SLIM (BSLIM) (shortly indicated as BIC-block sparse learning via iterative minimization (BSLIM)) procedure devised in [16] is used to initialize the algorithms and adopted as a term of comparison.

The considered simulating scenario comprises a spectrum sensing bandwidth of $B = 500$ MHz centered at the carrier frequency $f_0 = 2.4$ GHz. The sensor is equipped with a ULA with $M = 10$ equally spaced antennas, with interelement distance of $d = \lambda_0/2$. The first analyzed scenario assumes $K = 4$ emitters with their AOAs described by $\bar{\theta}_1 = 0°, \bar{\theta}_2 = 18°, \bar{\theta}_3 = 36°$, and $\bar{\theta}_4 = -18°$, and $\Delta\theta_1 = 0.643°$, $\Delta\theta_2 = 1.125°$, $\Delta\theta_3 = 0.750°$, and $\Delta\theta_4 = 0.900°$. Moreover, the first three emitters are communication sources operating on $[0.1B - \frac{1}{2T_s}(1+\beta), 0.1B + \frac{1}{2T_s}(1+\beta)]$, $[0.2B - \frac{1}{2T_s}(1+\beta), 0.2B + \frac{1}{2T_s}(1+\beta)]$, and $[\frac{1}{2T_s}(1+\beta), \frac{1}{2T_s}(1+\beta)]$, respectively, with $\beta = 0.5$ and $T_s = 10/B$. The fourth emitter is a jammer radiating a zero-mean circularly symmetric Gaussian signal with a flat spectrum over $[-0.1B, 0.1B]$. As to the communication sources, they transmit data via a quadrature phase shift keying modulation employing a root-raised-cosine pulse with roll-off parameter β and symbol rate T_s. Finally, $\sigma^2 = 0$ dB and two case studies are assessed: the former entails a signal-to-noise ratio (SNR) of 5 dB for all the emitters, the latter supposes SNR $= 10$ dB still for all the sources.

The discretization grid is realized with $K_1 = 40$ uniformly spaced angles over the interval $[-90°, 90°]$, corresponding to a step $\delta = 4.5°$. As to the available temporal observations, $N_1 = 10$ independent data-windows are processed each with $N = 50$ snapshots and $N_F = 2N$, i.e., twice the frequency resolution induced by the number of available samples. Regarding the parameter settings employed by Algorithm 2, $\xi = 10^{-1}$ and $n_{\max} = 30$ are used for the exit condition, a smoothing factor of $\epsilon = 10^{-6}$ is employed, the parameter q is fixed to 0.6, and the bounds on the interference power level are set as $\sigma_L^2 = -3$ dB and $\sigma_U^2 = 10$ dB. The same exit condition is applied for Algorithm 1 (viz., $\xi_1 = 10^{-1}$ and $n_{1_{\max}} = 30$). Moreover, the algorithms OG-MBI-OFDR and the OG-MBI-EFDR assume $\alpha = 0.05$ and $\kappa = 40$, whereas $\bar{K} = 180$ is adopted in OG-MBI-BIC.

Fig. 2 displays the nominal space-frequency profile of the aforementioned simulation scenario with SNR $= 10$ dB, obtained by evaluating for each angle bin the average (over the N_1 data-windows) energy spectral density of the signal received on the AOA of the associated emitter, normalized to the maximum value. The plot clearly demonstrates the space-frequency regions occupied by the four emitters. Moreover, being the power of the emitters equal, the larger

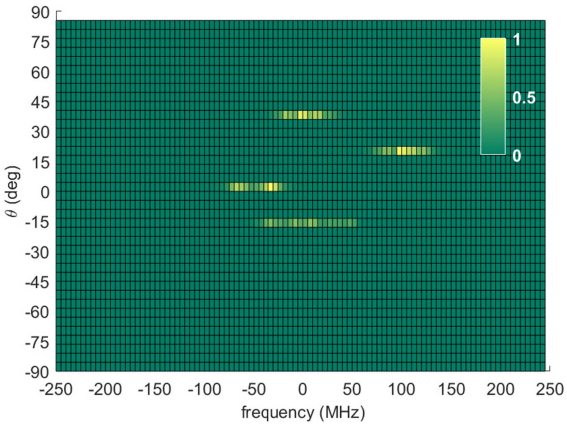


Fig. 2. Space-frequency profile of the first (off-grid) analyzed scenario with SNR $= 10$ dB.

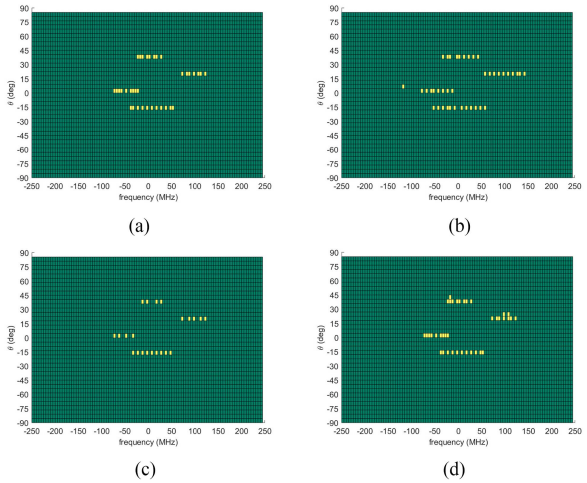


Fig. 3. Space-frequency occupancy map for the first (off-grid) analyzed scenario with SNR $= 10$ dB. Subplots refer to (a) OG-MBI-BIC, (b) OG-MBI-OFDR, (c) OG-MBI-EFDR, and (d) BIC-BSLIM [16].

the source bandwidth the smaller the Energy Spectral Density (ESD) values.

In Fig. 3, the results of the space-frequency occupancy maps recovery process provided by OG-MBI-BIC, OG-MBI-OFDR, OG-MBI-EFDR, and the BIC-BSLIM of [16] (for measurements collected at SNR $= 10$ dB) are illustrated, with reference to one simulation trial.

From a visual comparison of the reconstructed 2-D maps with the ground-truth, it is evident that both the OG-MBI-BIC and OG-MBI-OFDR techniques can provide an almost perfect recovery localizing the sources accurately in both the angle and frequency domains. However, the maps show that the OG-MBI-OFDR tends to spread the detection in the frequency domain, yielding to a higher number of false detections. Moreover, these two approaches outperform the OG-MBI-EFDR, which does not detect some active frequency bins. Nevertheless, all the novel off-grid recovery strategies localize, in the correct spatial bin, the four sources, without false alarms in other angular directions. On the other hand, BIC-BSLIM [16], which neglects the off-grid behavior, due to the model mismatches

TABLE II
Empirical False Alarm Rate, P_{fa}, and Empirical Detection Probability, P_d, for the Proposed OG-MBI-BIC, OG-MBI-OFDR and OG-MBI-EFDR Processing for the Sensing Scenario 1

	SNR $= 5$ dB		SNR $= 10$ dB	
	P_{fa}	P_d	P_{fa}	P_d
OG-MBI-BIC	~ 0	0.56	2.0×10^{-4}	0.94
OG-MBI-OFDR	28×10^{-4}	0.98	13×10^{-4}	0.98
OG-MBI-EFDR	~ 0	0.53	0.3×10^{-4}	0.79
BIC-BSLIM [16]	~ 0	0.55	8.2×10^{-4}	0.94

gives rise to false alarms at other AOAs that represent ghost sources. Noteworthy, all approaches detect only a subset of the frequency bins per emitter, providing maps exhibiting an on–off behavior, due to dictionary redundancy in the frequency domain and sparsity promoting nature of the developed procedures. Nevertheless, the results demonstrate that all the devised approaches are capable of obtaining a valid signal description.

To offer further insights and provide a global picture about the recovery capabilities of the proposed strategies, the behavior of the different strategies with respect to multiple trials is now analyzed. In this respect, the empirical false alarm rate P_{fa}, and the empirical detection probability P_d are used as performance metrics. In particular, empirical P_{fa} and P_d are obtained counting the number of detections in the source-free/occupied space-frequency bins, respectively. To handle the intrinsic on–off behavior of the recovered maps, a moving average filtering (with two equal weights) along the frequency dimension is performed over the recovered profile (containing in each space-frequency bin the estimated average energy), for all the considered techniques.[5]

In Table II, the empirical false alarm rate and detection probability are reported for the simulation setup of the scenario 1 illustrated in Fig. 2 and for two different SNR values, viz., SNR $= 5$ dB and SNR $= 10$ dB. Moreover, both P_{fa} and P_d are evaluated over ten independent trials, where at each run there are $N_1 = 10$ independent data-windows, corresponding to 39 370 emitters free space-frequency bins and 630 bins occupied by RF sources (for each communication emitter a guard cell accounting for the 3-dB bandwidth measure is also considered). The results clearly show that both the OG-MBI-BIC and the OG-MBI-OFDR can reach higher P_d values than the OG-MBI-EFDR, which instead has lower P_{fa} values. Notably, the OG-MBI-OFDR has the best detection capabilities, being able to ensure $P_d = 0.98$ also at low SNRs. This advantage is paid by its higher computational complexity, as also highlighted by the average execution times[6] reported in Table III. All the proposed off-grid methods for each source always detect at least one bin in correspondence to the associated spatial direction. Last but not least, the OG-MBI-BIC and OG-MBI-EFDR ensure lower P_{fa} values than the OG-MBI-OFDR. This

[5] Note that, other than the metrics used in this article, the false alarm probability can be also evaluated as in [64].

[6] The average execution time is evaluated as the mean time occupied over 10 Monte Carlo trials, run on an Intel Core i5-1135G7 at 2.40 GHz.

TABLE III
Average Execution Time of the Proposed Profile Refinement Strategies

refinement strategy	average execution time (s)
BIC based	0.0421 s
OMP-FDR based	0.6133 s
Energy profile-FDR based	0.0108 s

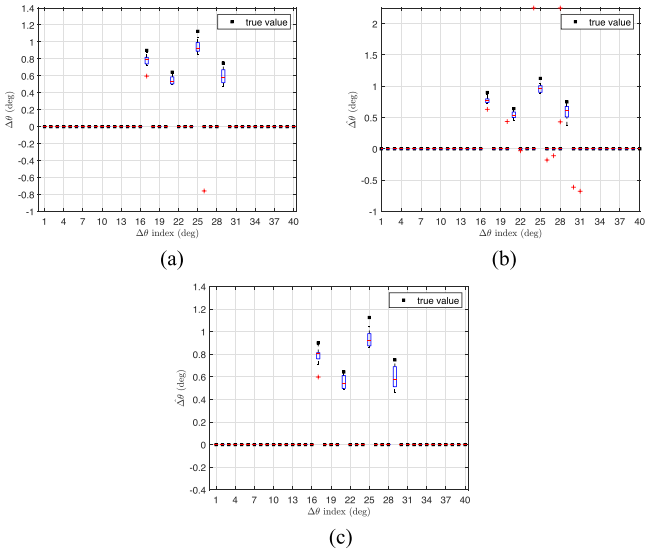


Fig. 4. Boxplots of the angle displacement estimation for the first analyzed scenario with SNR $= 10$ dB. Subplots refer to (a) OG-MBI-BIC, (b) OG-MBI-OFDR, and (c) OG-MBI-EFDR.

behavior is motivated by the fact that the OG-MBI-OFDR tends to provide an enlarged frequency support for each emitter. Finally, it is not surprising that the BIC-BSLIM is also capable of ensuring high detection performance, although it introduces false alarms in directions adjacent to those of the true sources, thereby resulting in ghost sources estimation.

To further shed light on the performance of the developed strategies, Fig. 4 reports the boxplots (with the central mark indicating the median, the edges denoting the 25th and 75th percentiles, the whiskers are extreme data points not considered outliers, and "+"-marker identifying outliers) of the angle displacement estimation for the first analyzed scenario with SNR $= 10$ dB. Subplots refer to 1) OG-MBI-BIC; 2) OG-MBI-OFDR; and 3) OG-MBI-EFDR. Inspection of the figures clearly unveils that accurate estimates of $\Delta\theta_i, i = 1, \ldots, K_1$, are obtained for all the considered approaches. Moreover, the OG-MBI-OFDR shows the presence of some outliers in the estimated values, whose number can be, however, controlled by properly tuning the parameter κ.

To quantitatively measure the accuracy of the AOA estimation of the proposed strategies, the root-mean-square error (RMSE) of the angle displacements is also evaluated. To this end, the focus is only on the bins occupied by the sources. In order to proceed, let $\widehat{\boldsymbol{\Delta\theta}}_r$ be the vector containing the displacement estimates associated with the source directions and let $\boldsymbol{\Delta\theta}_r$ be the vector whose entries are the corresponding true displacements. The RMSE (constrained

TABLE IV
RMSE of Angular Displacement (in Degrees) for the Proposed OG-MBI-BIC, OG-MBI-OFDR, and OG-MBI-EFDR Processing for the Sensing Scenario 1

	RMSE	
	SNR = 5 dB	SNR = 10 dB
OG-MBI-BIC	0.34 deg	0.31 deg
OG-MBI-OFDR	0.37 deg	0.32 deg
OG-MBI-EFDR	0.34 deg	0.33 deg
BIC-BSLIM [16]	1.75 deg	1.75 deg

to the actual source angle bins) is thus defined as

$$\text{RMSE} = \sqrt{\mathbb{E}\left[\left\|\widehat{\boldsymbol{\Delta\theta}}_r - \boldsymbol{\Delta\theta}_r\right\|^2\right]}. \tag{41}$$

Since, a closed-form expression for the RMSE in (41) is not available, the Monte Carlo simulation strategy is used for its computation. Table IV reports the RMSEs of all the considered algorithms with reference to the first analyzed scenario. In agreement with the results of Fig. 4, the diverse space-frequency recovery procedures are capable to accurately estimate the source AOA, i.e., correctly learn the actual model dictionary. It is not surprising that all the algorithms have approximately the same RMSE, outperforming the on-grid counterpart, which indeed does not account for any displacement in the considered grid.

As an additional term of comparison, the gridless method in [40] designed for 2-D DOA estimation is also considered. In particular, this method, denoted as 2D-MNOMP hereafter, involves the selection of some parameters, as detailed in the following (the related Matlab codes can be downloaded at https://github.com/RiverZhu/2DMNOMP). The oversampling factors (γ_x, γ_y) are chosen equal to (4, 4), and the model order overestimating probability is set as $P_{\text{oe}} = 10^{-4}$. Moreover, being the noise power σ^2 an unknown parameter (its true value is 0 dB), and only an estimate $\hat{\sigma}^2$ is reasonably available, the tests are conducted assuming either an ideal (i.e., perfect knowledge) matched case $\hat{\sigma}^2 = 0$ dB, or some mismatched (more practical) situations due to the unavoidable estimation errors (in particular, $\hat{\sigma}^2 = \{-3, 3\}$ dB). Results (not reported here for brevity) has shown that the 2D-MNOMP is capable of detecting the four sources, but its performance degrades as the mismatch in the noise power estimate increases. Specifically, when the noise power is assumed known (which is the setup with a less practical relevance), all the algorithms, i.e., both the proposed (with $\sigma_L^2 = \sigma_U^2 = 0$ dB) and the 2D-MNOMP, ensure very high detection capabilities, with a false alarm rate on the order of 10^{-4}. However, when the noise power is unknown, the proposed method still maintains nearly the same performance as in the ideal case. In contrast, the 2D-MNOMP tends to show a decrease in detection capabilities when the noise level is overestimated, and an increase in false alarms when it is underestimated. Regarding the RMSE values of the source angles,[7] the conducted experiments have highlighted that all the proposed methods and the 2D-MNOMP with known noise power achieve approximately the same performance. These results are confirmed also when an overestimation of the noise power is considered in the 2D-MNOMP algorithm, but at a cost of a reduced detection capability (as already stated). Moreover, the 2D-MNOMP exhibits a significant performance degradation, with an increasing RMSE in angle displacement estimation, when the noise level is underestimated.

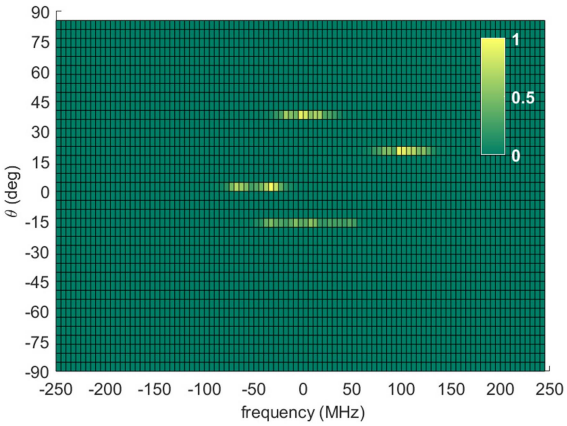


Fig. 5. Space-frequency profile of the second (on-grid) analyzed scenario with SNR = 10 dB.

Despite the results discussed above, it is worth to underline that the 2D-MNOMP performs harmonic retrieval with a uniform planar array, that is equivalent to the signal model considered for the case studies herein reported. However, this steering vectors structure is only a special case for the proposed framework; in fact, the designed algorithm is not restricted to estimating sinusoidal waves, but it can account for more general array manifolds possibly due to mutual coupling between array elements, presence of calibration errors, nonplanar and nonconformal array architectures, in addition, of course, to the spectral domain. Moreover, the proposed methodology can be extended to encompass sensors capable of acquiring polarimetric features of the impinging waves, so as to further boost surveillance capabilities [65], [66]. Finally, another advantage of the proposed method is that it can be extended to also account for the presence of missing data [67].

To grasp further insights on the potentiality of the proposed 2-D sensing approaches and evaluate its robustness, let us now focus on an on-grid scenario. Specifically, the same setup as the previous case study is considered, but with the sources AOAs lying on the grid, viz., $\boldsymbol{\Delta\theta} = \mathbf{0}$. Hence, the nominal space-frequency profile for SNR = 10 dB is shown in Fig. 5, where, as before, it is obtained evaluating for each angle bin the normalized average ESD of the signal

[7] For the 2D-MNOMP algorithm, each angle estimate (for a given spatial bin) is obtained by averaging the angles inferred from the different active space-frequency bins.

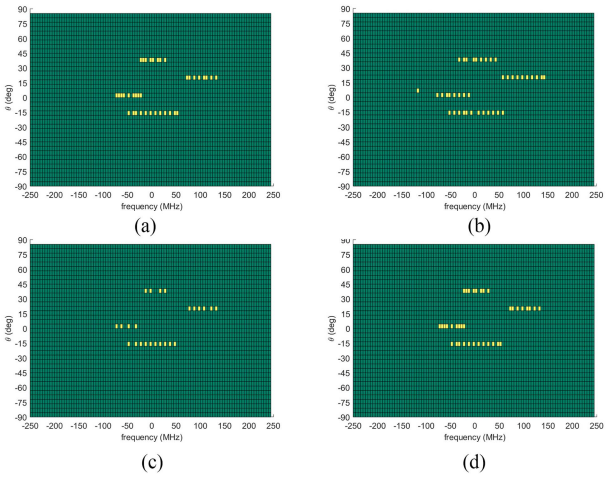


Fig. 6. Space-frequency occupancy map for the second (on-grid) analyzed scenario with SNR = 10 dB. Subplots refer to (a) OG-MBI-BIC, (b) OG-MBI-OFDR, (c) OG-MBI-EFDR, and (d) BIC-BSLIM [16]. (a) OG-MBI-BIC. (b) OG-MBI-OFDR. (c) OG-MBI-EFDR. (d) BIC-BSLIM.

TABLE V
Empirical False Alarm Rate, P_{fa}, and Empirical Detection Probability, P_d, for the Proposed OG-MBI-BIC, OG-MBI-OFDR and OG-MBI-EFDR Processing for the Sensing Scenario 2

	SNR = 5 dB		SNR = 10 dB	
	P_{fa}	P_d	P_{fa}	P_d
OG-MBI-BIC	~ 0	0.65	1.8×10^{-4}	0.96
OG-MBI-OFDR	35×10^{-4}	0.98	28×10^{-4}	0.98
OG-MBI-EFDR	~ 0	0.59	0.3×10^{-4}	0.76
BIC-BSLIM [16]	~ 0	0.66	1.78×10^{-4}	0.96

received on the considered AOA. It is noteworthy that due to the identical grid resolution in both scenarios, the maps depicted in Figs. 2 and 5 coincide.

Fig. 6 displays the space-frequency occupancy maps recovered via the proposed OG-MBI-BIC, OG-MBI-OFDR, OG-MBI-EFDR, as well as the BIC-BSLIM of [16], for a specific simulation trial. Comparing these figures emerges the fact that OG-MBI-BIC is capable of providing a profile reconstruction equal to that of the BIC-BSLIM specifically designed for the on-grid case. Differently, both the OG-MBI-OFDR and OG-MBI-EFDR experience some slight performance degradations. In particular, OG-MBI-OFDR shows some false alarms together with an excessive expansion of the reconstructed sources in the frequency domain, whereas OG-MBI-EFDR presents some additional miss detections in the source profile estimates. Nonetheless, all the provided off-grid strategies accomplish a reliable determination of the space-frequency activation profile, also in on-grid situations.

To further investigate the performance of the devised strategies, Table V summarize the empirical P_{fa} and P_d obtained by the proposed algorithms for the considered on-grid scenario. The obtained P_d and P_{fa} values confirm the capabilities of the proposed method to correctly detect the space-frequency profile. In particular, the OG-MBI-BIC share the same performance as the BIC-BSLIM. Moreover,

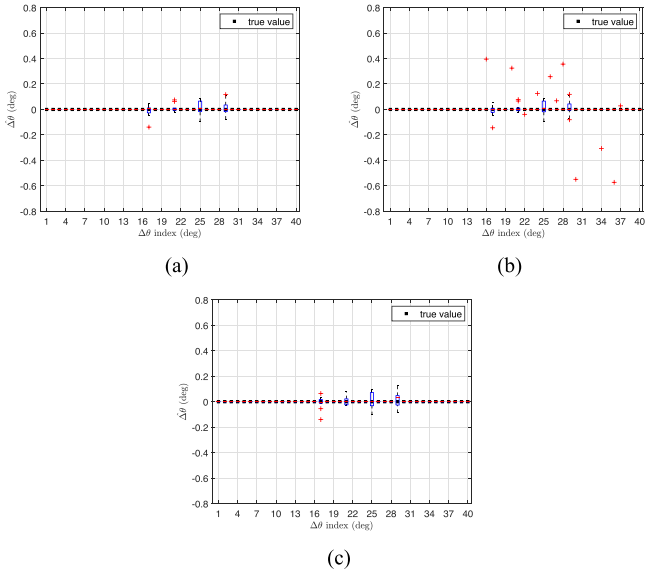


Fig. 7. Boxplots of the angle displacement estimation for the second analyzed scenario with SNR = 10 dB. Subplots refer to (a) OG-MBI-BIC, (b) OG-MBI-OFDR, and (c) OG-MBI-EFDR. (a) OG-MBI-BIC. (b) OG-MBI-OFDR. (c) OG-MBI-EFDR.

TABLE VI
RMSE of Angular Displacement (in Degrees) for the Proposed OG-MBI-BIC, OG-MBI-OFDR, and OG-MBI-EFDR Processing for the Sensing Scenario 2

	RMSE	
	SNR = 5	SNR = 10 dB
OG-MBI-BIC	0.16 deg	0.10 deg
OG-MBI-OFDR	0.15 deg	0.11 deg
OG-MBI-EFDR	0.17 deg	0.11 deg
BIC-BSLIM [16]	0 deg	0 deg

the OG-MBI-OFDR is still capable of outperforming the BIC-BSLIM in terms of number of detections, but at the expense of a higher number of false alarms. In this particular circumstance, the OG-MBI-EFDR exhibits the worst detection capabilities, even if all sources are detected. This behavior allows the OG-MBI-EFDR to significantly reduce the number of false alarms in the map.

To analyze the effectiveness of the proposed strategy to adequately infer the sources AOAs, Fig. 7 reports the boxplots of the angle displacement estimation for the second (on-grid) analyzed scenario with SNR = 10 dB. Subplots refer to 1) OG-MBI-BIC; 2) OG-MBI-OFDR; and 3) OG-MBI-EFDR. Notably, while BIC-BSLIM operates in matched conditions, the proposed off-grid strategies do not. In particular, all of them have to estimate a higher number of parameters (i.e., the angular displacements) leading to some performance degradations. However, the graphs show that the estimates of $\Delta\theta_i$, $i = 1, \ldots, K_1$, are strongly concentrated around zero with a slight dispersion in correspondence to the actual source AOAs. The boxplots also highlight that the OG-MBI-OFDR generates few outliers in some free-source directions. Noteworthy, the respective error, quantified by the RMSE, are less than in the previous analysis, as specified in Table VI.

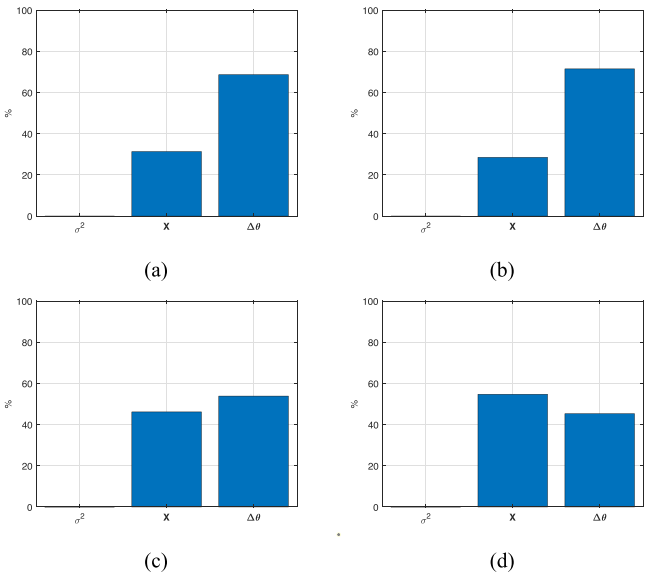


Fig. 8. Optimization results for the two considered environmental scenarios in terms of percentage of MBI selections. Subplots refer to (a) scenario 1 with SNR = 5 dB, (b) scenario 1 with SNR = 10 dB, (c) scenario 2 with SNR = 5 dB, and (d) scenario 2 with SNR = 10 dB.

Before concluding this section, in Fig. 8 the percent of times each variable block is optimized by Algorithm 2 for all Monte Carlo runs is displayed. The inspection of the figures clearly reveals that a major role in the inference process is played by the angle displacement update, i.e., the dictionary learning process. It is also worth pointing out that, the noise level is never updated, since its initial estimate is very close to its true value (e.g., for the scenario 1 with SNR = 10 dB, its average estimate on the Monte Carlo trials is equal to 1.04 with a dispersion around the mean value of 0.05).

VI. CONCLUSION

This article has considered the problem of off-grid 2-D spectrum sensing for cognitive radars. Precisely, an off-grid 2-D profile recovery strategy, framed as an RML estimation, aimed at leveraging the inherent block-sparsity of the entire profile, is proposed. The resulting optimization problem was addressed using an MM-MBI approach, which iteratively refines the estimates of three variable blocks, viz., noise power, 2-D profile, and angular displacements. Successively, to improve the reliability of the constructed space-frequency occupancy map and to get a more accurate estimate of the AOAs, three refinement procedures based on the selection of the most dominant active atoms in the space-frequency profile have been also conceived. Numerical simulations have demonstrated the validity of the proposed framework, with the BIC-based refinement strategy showing the best performance in terms of tradeoff between computational complexity and estimation accuracy.

Summarizing, the main technical contributions provided in this article are as follows.

1) The formulation of the space-frequency profile recovery problem as a solution to the RML estimation, with a joint optimization of the noise power, 2-D profile, and angular displacements.
2) The design of a novel algorithm based on MM-MBI to address the challenging optimization problem. Precisely, it involves the solution of three subproblems, one for each of the three variable blocks, that are optimally solved. For each iteration, the MBI updates the variable block leading to the larger decrement of the objective function, while keeping the other blocks fixed.
3) The analysis of the convergence properties of the proposed algorithm, showing that any cluster point of the produced sequence of estimates fulfills the KKT conditions.
4) The development of three refinement strategies aimed at improving the reliability of the space-frequency occupancy map recovery process along with the accuracy of the angle displacements.
5) The assessment of the effectiveness of the novel sensing procedures also with reference to their computational complexity.

Future works would consider the extension to both azimuth and elevation angles, as well as testing the devised framework on measured data. The possibility of performing a bespoke adaptive Taylor expansion to define appropriate (i.e., with convergence guarantees) Newton-like methods is also an area for future research.

APPENDIX A SOLUTION TO PROBLEM (28)

In order to determine the closed-form solution for a specific displacement angle, i.e., for a given $i \in \{1, \ldots, K_1\}$, to Problem (28), the objective function of $\mathcal{P}_3$ is now appropriately reformulated. To this end, based on (12),

$$\boldsymbol{H}(\boldsymbol{\Delta\theta}) = \bar{\boldsymbol{H}} + \dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)$$

with

$$\dot{\boldsymbol{H}} = \left[\dot{\boldsymbol{s}}(\theta_1, \omega_1), \ldots, \dot{\boldsymbol{s}}(\theta_1, \omega_{N_F}), \dot{\boldsymbol{s}}(\theta_2, \omega_1), \ldots, \dot{\boldsymbol{s}}(\theta_{K_1}, \omega_{N_F})\right] \in \mathbb{C}^{MN \times K_1 N_F}. \tag{42}$$

As a consequence,

$$\boldsymbol{H}(\boldsymbol{\Delta\theta})\boldsymbol{X} = \bar{\boldsymbol{H}}\boldsymbol{X} + \dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)\boldsymbol{X}. \tag{43}$$

To proceed further, let us observe that

$$\begin{aligned}
&\dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)\boldsymbol{X} \\
&= \left[\dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)\boldsymbol{x}^1, \ldots, \dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)\boldsymbol{x}^{N_1}\right] \\
&= \left[\dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{x}^1\right)\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right), \ldots, \dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{x}^{N_1}\right)\right. \\
&\quad \left.\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)\right].
\end{aligned} \tag{44}$$

Hence, the objective function of Problem $\mathcal{P}_3$ can be expressed as

$$\begin{aligned}
&\left\|\bar{\boldsymbol{H}}\boldsymbol{X} + \dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right)\boldsymbol{X} - \boldsymbol{Y}\right\|^2 \\
&= \sum_{i=1}^{N_1}\left\|\dot{\boldsymbol{H}}\mathbf{diag}\left(\boldsymbol{x}^i\right)\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right) - \boldsymbol{z}^i\right\|^2 \qquad (45) \\
&= \left\|\boldsymbol{B}\left(\boldsymbol{\Delta\theta} \otimes \mathbf{1}_{N_F}\right) - \boldsymbol{\zeta}\right\|^2 \qquad (46)
\end{aligned}$$

where

$$\boldsymbol{Z} = [z^1, \ldots, z^{N_1}] = \boldsymbol{Y} - \bar{\boldsymbol{H}}\boldsymbol{X}, \tag{47}$$

$$\boldsymbol{B} = \begin{bmatrix} \dot{\boldsymbol{H}}\,\mathbf{diag}\,(\boldsymbol{x}^1) \\ \vdots \\ \dot{\boldsymbol{H}}\,\mathbf{diag}\,(\boldsymbol{x}^{N_1}) \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\zeta} = \begin{bmatrix} z^1 \\ \vdots \\ z^{N_1} \end{bmatrix}. \tag{48}$$

Finally, the function in (46) can be equivalently cast as

$$\|\boldsymbol{C}\,\Delta\boldsymbol{\theta} - \boldsymbol{\zeta}\|^2 \tag{49}$$

where the ith column of $\boldsymbol{C} \in \mathbb{C}^{MNN_1 \times K_1}$ is[8]

$$\boldsymbol{c}_i = \sum_{k=1}^{N_F} \boldsymbol{B}_{(i-1)N_F+k}. \tag{50}$$

Leveraging (49), Problem $\mathcal{P}_3^i$, $i = 1, \ldots, K_1$, (to be solved in the inner loop of Algorithm 1) can be written as the following constrained minimization problem:

$$\begin{cases} \min\limits_{\Delta\theta_i} & \|\boldsymbol{c}_i\Delta\theta_i - \boldsymbol{\zeta} + \boldsymbol{C}_{-i}\Delta\boldsymbol{\theta}_{-i}\|^2 \\ \text{s.t.} & |\Delta\theta_i| \le \delta/2 \end{cases} \tag{51}$$

or equivalently

$$\begin{cases} \min\limits_{\Delta\theta_i} & \|\boldsymbol{c}_i\|^2(\Delta\theta_i)^2 + \|\boldsymbol{\zeta} - \boldsymbol{C}_{-i}\Delta\boldsymbol{\theta}_{-i}\|^2 \\ & \quad -2\Delta\theta_i \mathrm{Re}\left[\boldsymbol{c}_i{}^{\dagger}\left(\boldsymbol{\zeta} - \boldsymbol{C}_{-i}\Delta\boldsymbol{\theta}_{-i}\right)\right] \\ \text{s.t.} & |\Delta\theta_i| \le \delta/2. \end{cases} \tag{52}$$

Now, if $\|\boldsymbol{c}_i\| > 0$, (52) is a strictly convex problem [60], whose global optimal solution can be obtained by setting equal to zero its first derivative and solving for $\Delta\theta_i$, leading to

$$\Delta\theta_i = \min\left(\frac{\delta}{2}, \max\left(-\frac{\delta}{2}, \frac{\mathrm{Re}\left[\boldsymbol{c}_i{}^{\dagger}\left(\boldsymbol{\zeta} - \boldsymbol{C}_{-i}\Delta\boldsymbol{\theta}_{-i}\right)\right]}{\|\boldsymbol{c}_i\|^2}\right)\right). \tag{53}$$

Otherwise, any feasible point is optimal, in particular, that described by (53) which implies a zero displacement.

APPENDIX B PROOF OF PROPOSITION 1

As to claim 1) of Proposition 1, being

$$\begin{aligned} g\left(\sigma^2, \boldsymbol{X}, \Delta\boldsymbol{\theta}\right) &\ge \log\sigma^2 + f_1(\boldsymbol{X}) \\ &\forall\, \sigma_L^2 \le \sigma^2 \le \sigma_U^2 \\ &\forall\, \boldsymbol{X} \in \mathbb{C}^{K_1N_F \times N_1} \\ &\forall\, \Delta\theta : |\Delta\theta_h| \le \delta/2,\ h = 1, \ldots, K_1 \end{aligned} \tag{54}$$

it follows that

$$g\left(\sigma^2, \boldsymbol{X}, \Delta\boldsymbol{\theta}\right) \to \infty \quad \text{as} \quad \|\boldsymbol{X}\|^2 \to \infty.$$

[8] If $\boldsymbol{C}$ is a full-rank column matrix, Problem $\mathcal{P}_3$ is strictly convex and Algorithm 1 provides the global optimal solution.

As a consequence, there exists a constant $K_X > 0$ such that Problem (15) is equivalent to

$$\begin{cases} \min\limits_{\boldsymbol{X}, \sigma^2, \Delta\boldsymbol{\theta}} & NMN_1 \log(\sigma^2) \\ & \quad + \frac{1}{\sigma^2}\|\boldsymbol{H}(\Delta\boldsymbol{\theta})\boldsymbol{X} - \boldsymbol{Y}\|^2 + f_1(\boldsymbol{X}) \\ \text{s.t.} & \sigma_L^2 \le \sigma^2 \le \sigma_U^2 \\ & |\Delta\theta_h| \le \delta/2,\ h = 1, \ldots, K_1 \\ & \|\boldsymbol{X}\|^2 \le K_X. \end{cases} \tag{55}$$

Now, since the feasible set of (55) is compact and the objective a continuous function, according to the Weierstrass theorem [68], it admits a global optimal solution. As a consequence $\nu(\mathcal{P}) \le g(\sigma^2, \boldsymbol{X}, \Delta\boldsymbol{\theta})$ for all $\sigma_L^2 \le \sigma^2 \le \sigma_U^2$, $\boldsymbol{X} \in \mathbb{C}^{K_1N_F \times N_1}$, and $\Delta\theta$ such that $|\Delta\theta_h| \le \delta/2$, $h = 1, \ldots, K_1$.

Moreover, $\sigma^{2(n)}$ and $\Delta\boldsymbol{\theta}^{(n)}$ are bounded by design, whereas

$$\begin{aligned} \|\boldsymbol{X}^{(n)}\|^2 = & \\ & \mathrm{tr}\left\{\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)})\left(\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\right.\right. \\ & \left.\left.\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)}) + \sigma^{2(n-1)}\boldsymbol{I}\right)^{-1}\boldsymbol{Y}\boldsymbol{Y}^{\dagger} \right. \\ & \left.\left(\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)}) + \sigma^{2(n-1)}\boldsymbol{I}\right)^{-1}\right. \\ & \left.\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\right\} \\ & \overset{(a)}{\le} \lambda_{\max}\left(\boldsymbol{Y}\boldsymbol{Y}^{\dagger}\right)\mathrm{tr}\left\{\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)})\right. \\ & \left.\left(\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)}) + \sigma^{2(n-1)}\boldsymbol{I}\right)^{-1}\right. \\ & \left.\left(\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)}) + \sigma^{2(n-1)}\boldsymbol{I}\right)^{-1}\right. \\ & \left.\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\right\} \\ & \overset{(b)}{\le} \frac{\lambda_{\max}\left(\boldsymbol{Y}\boldsymbol{Y}^{\dagger}\right)}{\left(\sigma_L^2\right)^2}\mathrm{tr}\left\{\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)})\right. \\ & \left.\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{D}_{\boldsymbol{X}^{(n-1)}}^{-2}\right\} \\ & \overset{(c)}{\le} \frac{\lambda_{\max}\left(\boldsymbol{Y}\boldsymbol{Y}^{\dagger}\right)}{\left(\sigma_L^2\right)^2}\left((K_x+\epsilon)^{1-\frac{q}{2}}\right)^2 \\ & \mathrm{tr}\left\{\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta}^{(n-1)})\boldsymbol{H}(\Delta\boldsymbol{\theta}^{(n-1)})\right\} \\ & \overset{(d)}{\le} \frac{\lambda_{\max}\left(\boldsymbol{Y}\boldsymbol{Y}^{\dagger}\right)}{\left(\sigma_L^2\right)^2}(K_x+1)^2 M_H \end{aligned} \tag{56}$$

where (a)–(c) in (56) stem from

$$\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{\dagger}\right) \le \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{B}_1\boldsymbol{A}^{\dagger}\right) \text{ if } \boldsymbol{B} \preceq \boldsymbol{B}_1$$

whereas (d) follows from

$$\begin{aligned} \left((K_x+\epsilon)^{1-\frac{q}{2}}\right)^2 \le (K_x+1)^{2-q} \le (K_x+1)^2 \\ \forall\, 0 \le \epsilon \le 1\ \forall\, q \in [0, 2] \end{aligned} \tag{57}$$

and

$$M_H = \max_{\substack{|\Delta\theta_h| \le \delta/2 \\ h=1,\ldots,K_1}} \mathrm{tr}\left\{\boldsymbol{H}^{\dagger}(\Delta\boldsymbol{\theta})\boldsymbol{H}(\Delta\boldsymbol{\theta})\right\}$$

which is a finite value due to the Weirstrass theorem.

As to the monotonicity and convergence of $g(\mathcal{Y}^{(n)})$, i.e., claim 2) of Proposition 1, they are a direct consequence of the block MM-MBI process [56], [57], the features of Algorithm 2, and the solvability of $\mathcal{P}$. Moreover, exploiting the continuity of the objective function, for any cluster point $\mathcal{Y}^{(\star)}$ of $\mathcal{Y}^{(n)}$, let us extract a converging subsequence $\mathcal{Y}^{(n_1)}$ to $\mathcal{Y}^{(\star)}$; then, due to the continuity of $g(\mathcal{Y})$ it follows

$$g^* = \lim_{n\to\infty} g\left(\mathcal{Y}^{(n)}\right) = \lim_{n\to\infty} g\left(\mathcal{Y}^{(n_1)}\right) = g\left(\mathcal{Y}^{(\star)}\right). \tag{58}$$

As to claim 3) of Proposition 1, the final step of the proof requires establishing that any limit point $\mathcal{Y}^{(\star)}$ fulfills KKT conditions for Problem $\mathcal{P}$. Before proceeding further, let us observe that the objective function of Problem $\mathcal{P}$ is continuously differentiable and Problem $\mathcal{P}$ is equivalent to Problem (55) with K_X such that

$$g(\mathcal{Y}) > g\left(\mathcal{Y}^{(0)}\right) \tag{59}$$

for all $\sigma_L^2 \le \sigma^2 \le \sigma_U^2$, for all $\Delta\theta_h \in [\delta/2, \delta/2]$, $h = 1, \ldots, K_1$, and for all $\boldsymbol{X}$ such that $\|\boldsymbol{X}\|^2 \ge K_X$.

Moreover, for each block of variables, the corresponding constraints satisfy linear independence constraint qualification [68] conditions. Indeed, for the cases of angle displacements and noise variance the constraints are described by affine functions, and there is at most one active constraint per variable of each block. Moreover, for the case of complex space-frequency profile, it is involved a nonactive quadratic constraint.

Hence, being the constraint functions separable w.r.t. each block variable and invoking Proposition 2 of [57], it follows that any limit point $\mathcal{Y}^{(\star)}$ satisfies the KKT conditions for Problem $\mathcal{P}$.

REFERENCES

[1] A. Farina, A. De Maio, and S. Haykin, *The Impact of Cognition on Radar Technology*. Rijeka, Croatia: Scitech, 2017.

[2] M. Wicks, "Spectrum crowding and cognitive radar," in *Proc. 2nd Int. Workshop Cogn. Inf. Process.*, 2010, pp. 452–457.

[3] M. A. Govoni, "Enhancing spectrum coexistence using radar waveform diversity," in *Proc. IEEE Radar Conf.*, 2016, pp. 1–5.

[4] H. He, P. Stoica, and J. Li, "Waveform design with stopband and correlation constraints for cognitive radar," in *Proc. 2nd Int. Workshop Cogn. Inf. Process.*, 2010, pp. 344–349.

[5] A. Aubry, A. De Maio, M. Piezzo, M. M. Naghsh, M. Soltanalian, and P. Stoica, "Cognitive radar waveform design for spectral coexistence in signal-dependent interference," in *Proc. IEEE Radar Conf.*, 2014, pp. 474–478.

[6] A. Aubry, V. Carotenuto, A. De Maio, A. Farina, and L. Pallotta, "Optimization theory-based radar waveform design for spectrally dense environments," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 31, no. 12, pp. 14–25, Dec. 2016.

[7] H. Griffiths et al., "Radar spectrum engineering and management: Technical and regulatory issues," *Proc. IEEE*, vol. 103, no. 1, pp. 85–102, Jan. 2015.

[8] Y. Zhao, J. Gaeddert, K. K. Bae, and J. H. Reed, "Radio environment map enabled situation-aware cognitive radio learning algorithms," in *Proc. Softw. Defined Radio Forum Tech. Conf.*, 2006, pp. 1–6.

[9] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in *Proc. 1st IEEE Int. Symp. New Front. Dyn. Spectr. Access Netw.*, 2005, pp. 151–159.

[10] F. Benedetto, G. Giunta, and L. Pallotta, "Cognitive satellite communications spectrum sensing based on higher order moments," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 574–578, Feb. 2021.

[11] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tut.*, vol. 11, no. 1, pp. 116–130, Firstquarter 2009.

[12] A. De Maio, Y. C. Eldar, and A. M. Haimovich, *Compressed Sensing in Radar Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[13] D. A. Guimaraes, R. A. A. de Souza, and A. N. Barreto, "Performance of cooperative eigenvalue spectrum sensing with a realistic receiver model under impulsive noise," *J. Sensor Actuator Netw.*, vol. 2, pp. 46–69, 2013.

[14] A. Fehske, J. Gaeddert, and J. Reed, "A new approach to signal classification using spectral correlation and neural netw.," in *Proc. 1st IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, Baltimore, MD, USA, 2005, pp. 144–150.

[15] A. Aubry, A. De Maio, and M. Govoni, "Two-dimensional spectrum sensing for cognitive radar," in *Proc. IEEE Radar Conf.*, 2018, pp. 815–820.

[16] A. Aubry, V. Carotenuto, A. De Maio, and M. A. Govoni, "Multi-snapshot spectrum sensing for cognitive radar via block-sparsity exploitation," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1396–1406, Mar. 2019.

[17] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse learning via iterative minimization with application to MIMO radar imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1088–1101, Mar. 2011.

[18] S. Han, L. Pallotta, X. Huang, G. Giunta, and D. Orlando, "A sparse learning approach to the design of radar tunable architectures with enhanced selectivity properties," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 3840–3853, Oct. 2020.

[19] L. Yan, P. Addabbo, C. Hao, D. Orlando, and J. Liu, "A sparse learning approach to multiple noise-like Jammers detection," in *Proc. Photon. Electromagnetics Res. Symp.-Fall*, 2019, pp. 155–161.

[20] A. Aubry, V. Carotenuto, A. De Maio, M. A. Govoni, and A. Farina, "Experimental analysis of block-sparsity-based spectrum sensing techniques for cognitive radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 1, pp. 355–370, Feb. 2021.

[21] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*, vol. 7. Amsterdam, The Netherlands: Elsevier, 2018, pp. 509–581.

[22] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[23] J. Zheng and M. Kaveh, "Directions-of-arrival estimation using a sparse spatial spectrum model with uncertainty," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 2848–2851.

[24] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4658–4671, Sep. 2012.

[25] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[26] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatch," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4997–5008, Oct. 2014.

[27] Q. Liu, H. C. So, and Y. Gu, "Off-grid DOA estimation with nonconvex regularization via joint sparse representation," *Signal Process.*, vol. 140, pp. 171–176, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168417301913

[28] L. Pallotta, G. Giunta, and A. Farina, "DOA refinement through complex parabolic interpolation of a sparse recovered signal," *IEEE Signal Process. Lett.*, vol. 28, pp. 274–278, 2021.

[29] M. Ibrahim, F. Romer, R. Alieiev, G. Del Galdo, and R. S. Thoma, "On the estimation of grid offsets in CS-based direction-of-arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 6776–6780.

[30] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3809–3822, Jul. 2012.
[31] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
[32] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
[33] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb. 2016.
[34] M. Sánchez-Fernández, V. Jamali, J. Llorca, and A. M. Tulino, "Gridless multidimensional angle-of-arrival estimation for arbitrary 3D antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4748–4764, Jul. 2021.
[35] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.
[36] A. Farina, *Antenna-Based Signal Processing Techniques for Radar Systems*. Norwood, MA, USA: Artech House, 1992.
[37] J. Shen, F. Gini, M. Greco, and T. Zhou, "Off-grid DOA estimation using improved root sparse Bayesian learning for non-uniform linear arrays," *EURASIP J. Adv. Signal Process.*, vol. 2023, 2023, Art. no. 34.
[38] Y. Wu, M. B. Wakin, and P. Gerstoft, "Gridless DOA estimation with multiple frequencies," *IEEE Trans. Signal Process.*, vol. 71, pp. 417–432, 2023.
[39] J. Zhu, L. Han, R. S. Blum, and Z. Xu, "Multi-snapshot newtonized orthogonal matching pursuit for line spectrum estimation with multiple measurement vectors," *Signal Process.*, vol. 165, pp. 175–185, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168419302658
[40] L. Han, X. Liu, N. Zhang, S. Wu, J. Zhu, and Z. Xu, "Two-dimensional multi-snapshot newtonized orthogonal matching pursuit for DOA estimation," *Digit. Signal Process.*, vol. 121, 2022, Art. no. 103313.
[41] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[42] P. Stoica and P. Babu, "False discovery rate (FDR) and familywise error rate (FER) rules for model selection in signal processing applications," *IEEE Open J. Signal Process.*, vol. 3, pp. 403–416, 2022.
[43] P. Babu and P. Stoica, "Multiple-hypothesis testing rules for high-dimensional model selection and sparse-parameter estimation," *Signal Process.*, vol. 213, 2023, Art. no. 109189.
[44] A. Aubry, A. De Maio, S. Marano, and M. Rosamilia, "Single-pulse simultaneous target detection and angle estimation in a multi-channel phased array radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 6649–6664, 2020.
[45] L. L. Scharf, *Statistical Signal Processing*. Reading, MA, USA: Addison-Wesley, 1991.
[46] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, 1st ed, vol. 2. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.
[47] P. Stoica and R. L. Moses, Eds., *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Pearson Prentice-Hall, 2006.
[48] S. D. Blunt, K. Gerlach, and T. Higgins, "Aspects of radar range super-resolution," in *Proc. IEEE Radar Conf.*, 2007, pp. 683–687.
[49] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*. Rijeka, Croatia: SciTech, 2010.
[50] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. A. Sadjadi, "Iterative adaptive approaches to MIMO radar imaging," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 5–20, Feb. 2010.
[51] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
[52] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
[53] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[54] Y. Wang, J. Wang, and Z. Xu, "On recovery of block-sparse signals via mixed l_2/l_q ($0 < q \leq 1$) norm minimization," *EURASIP J. Adv. Signal Process.*, vol. 2013, 2013, Art. no. 76.
[55] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Maximum likelihood estimation of a structured covariance matrix with a condition number constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, Jun. 2012.
[56] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
[57] A. Aubry, A. De Maio, A. Zappone, M. Razaviyayn, and Z.-Q. Luo, "A new sequential optimization procedure and its applications to resource allocation for wireless systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6518–6533, Dec. 2018.
[58] L. Lan, M. Rosamilia, A. Aubry, A. De Maio, and G. Liao, "FDA-MIMO transmitter and receiver optimization," *IEEE Trans. Signal Process.*, vol. 72, pp. 1576–1589, 2024.
[59] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.
[60] S. Boyd, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[61] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*. Burlington, MA, USA: Elsevier Academic, 2005.
[62] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: MPS-SIAM, 2001.
[63] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.
[64] M. Xu, J. Zhu, J. Fang, N. Zhang, and Z. Xu, "CFAR based NOMP for line spectral estimation and detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 6971–6990, Oct. 2023.
[65] A. Aubry, A. De Maio, and A. Farina, *Polarimetric Radar Signal Processing*. Stevenage, U.K.: SciTech, 2022.
[66] A. Aubry, M. Boddi, A. De Maio, and M. Rosamilia, "Sparse DOA estimation with polarimetric arrays," *IEEE Open J. Signal Process.*, vol. 5, pp. 886–901, 2024.
[67] A. Aubry, A. De Maio, S. Marano, and M. Rosamilia, "Structured covariance matrix estimation with missing-(Complex) data for radar applications via expectation-maximization," *IEEE Trans. Signal Process.*, vol. 69, pp. 5920–5934, 2021.
[68] D. P. Bertsekas, "Nonlinear programming," *J. Oper. Res. Soc.*, vol. 48, no. 3, pp. 334–334, 1997.

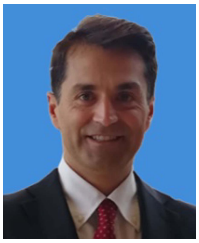
Augusto Aubry (Senior Member, IEEE) received the Dr. Eng. degree in telecommunication engineering (with honors) and the Ph.D. degree in electronic and telecommunication engineering both from the University of Naples Federico II, Naples, Italy, in 2007 and 2011, respectively.

From February to April 2012, he was a Visiting Researcher with the Hong Kong Baptist University, Hong Kong. He is currently an Associate Professor with the University of Naples Federico II. His research interests include statistical signal processing and optimization theory, with emphasis on MIMO communications and radar signal processing.

Dr. Aubry was the co-recipient of the 2013 Best Paper Award (entitled to B. Carlton) of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS with the contribution "Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter," and also the recipient of the 2022 IEEE Fred Nathanson Memorial Award as the young (less than 40 years of age) AESS Radar Engineer 2022, with the following citation "For outstanding contributions to the application of modern optimization theory to radar waveform design and adaptive signal processing."

Prabhu Babu received the B. Tech degree in electronics from the Madras Institute of Technology, Chennai, Tamil Nadu, India, in 2005, the M. Tech degree in radio frequency design technology from Centre for Applied Research in Electronics (CARE), Indian Institute of Technology (IIT) Delhi, New Delhi, Delhi, India, in 2007, and the doctor of philosophy degree in signal processing from Uppsala University, Uppsala, Sweden, in 2012.

From 2013 to 2015, he was with the Hong Kong University of Science and Technology, Hong Kong, doing his postdoctoral research. In 2016, he joined CARE, IIT Delhi, where currently he is an Associate Professor.



Antonio De Maio (Fellow, IEEE) received the Dr. Eng. (Hons.) and Ph.D. degrees in information engineering from the University of Naples Federico II, Naples, Italy, in 1998 and 2002, respectively.

From October to December 2004, he was a Visiting Researcher with the U.S. Air Force Research Laboratory, Rome, NY, USA. From November to December 2007, he was a Visiting Researcher with the Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the University of Naples Federico II. His research interests include the field of statistical signal processing, with emphasis on radar detection, optimization theory applied to radar signal processing, and multiple-access communications.

Dr. De Maio was the recipient of the 2010 IEEE Fred Nathanson Memorial Award as the young (less than 40 years of age) AESS Radar Engineer 2010 whose performance is particularly noteworthy as evidenced by contributions to the radar art over a period of several years, with the following citation for "robust CFAR detection, knowledge-based radar signal processing, and waveform design and diversity." He was also the corecipient of the 2013 best paper award (entitled to B. Carlton) of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS with the contribution "Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter." He was also a recipient of the 2024 IEEE Warren White Award for outstanding achievements due to a major technical advance (or series of advances) in the art of radar engineering, with the citation "For Contributions to Radar Signal Processing Techniques for Target Detection, Waveform Design, and Electronic Protection."

Luca Pallotta (Senior Member, IEEE) received the Laurea Specialistica degree (cum laude) in telecommunication engineering from the University of Sannio, Benevento, Italy, in 2009, and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, Italy, in 2014.

He was an Assistant Professor with the University of Roma Tre, Rome, Italy, from 2019 to 2022. He is currently an Assistant Professor with University of Basilicata, Potenza, Italy. His research interest lies in the field of statistical signal processing, with emphasis on radar/SAR signal processing, radar targets detection and classification, polarimetric radar/SAR.

Dr. Pallotta won the Student Paper Competition at IEEE Radar Conference 2013. He has been an Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2020. From 2018 to 2021, he was an Associate Editor for the Springer journal Signal, Image and Video Processing.