# Off-Grid Direction-of-Arrival Estimation Using Second-Order Taylor Approximation

Huiping Huang<sup>a</sup>, Hing Cheung So<sup>b</sup>, Abdelhak M. Zoubir<sup>a</sup>

<sup>a</sup>Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt, Darmstadt, Germany <sup>b</sup>Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China

# Abstract

The problem of off-grid direction-of-arrival (DOA) estimation is investigated. We develop a grid-based method to jointly estimate the closest *spatial frequency* (the sine of DOA) grids, and the gaps between the estimated grids and the corresponding frequencies. By using a second-order Taylor approximation, the data model under the framework of joint-sparse representation is formulated. We point out an important property of the signals of interest in the model, namely the proportionality relationship, which is empirically demonstrated to be useful in the sense that it increases the probability of the mixing matrix satisfying the block restricted isometry property. Simulation examples demonstrate the effectiveness and superiority of the proposed method against several state-of-the-art grid-based approaches.

*Keywords:* Block restricted isometry property (RIP), compressive sensing, off-grid DOA estimation, second-order Taylor approximation

## 1. Introduction

Grid-based methods have gained interest in direction-of-arrival (DOA) estimation in recent years. Such approaches include least absolute shrinkage and selection operator (LASSO) [1–3] and sparse iterative covariance-based estimation [4–7], among others. See [8] for a comprehensive review of grid-based sparse methods for DOA estimation. The advantage of grid-based methods is that they have super-high resolution even in the case when only one single snapshot is available, provided that all the source spatial frequencies align exactly with the preset grid. However, this condition may not be satisfied in practice, since the region of interest (ROI) contains infinite candidates and hence grid mismatch almost always exists when we split the ROI into a finite number of grids. This is known as the off-grid issue and has attracted a lot of research interest in array signal processing during the past decade, see for example [9–25].

Existing solutions to tackle the off-grid problem can be categorized into three groups. The first group uses denser grids or the coarse-to-fine strategy such as [2]. The drawbacks of these methods are twofold. On

<sup>\*</sup>The work of H. Huang is supported by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt.

*Email addresses:* h.huang@spg.tu-darmstadt.de (Huiping Huang), hcso@ee.cityu.edu.hk (Hing Cheung So), zoubir@spg.tu-darmstadt.de (Abdelhak M. Zoubir)

one hand, denser grids lead to extremely expensive computational complexity; on the other hand, too dense grids may result in weak incoherence among the steering vectors. The second group consists of the so-called *gridless* approach [26–31]. Its weakness is that most of these methods are restricted to regularly sampled measurements that can only be taken from a uniform linear array (ULA) [32]. The last group of methods estimates the off-grid bias together with the grids closest to the true spatial frequencies. Representative works include the first-order Taylor approximation [13, 15] and the neighbor-grid based method [19], denoted in this paper as 1st Taylor G-LASSO and Neighbor G-LASSO, respectively.

It is known that in general the first-order Taylor approximation is accurate enough, especially when the grid size is small. However, when the grid size is set not small enough so as to save computational cost, there still exists a large bias. In such a situation, a high-order Taylor approximation decreases the approximation error. To this end, we introduce a second-order Taylor approximation in off-grid DOA estimation. We observe in this case the proportionality relationship of the signals of interest. With this, we propose a novel optimization approach which is shown by simulation to produce more accurate frequency estimates in off-grid scenarios. Moreover, the uniqueness issue of the proposed method is discussed by means of the restricted isometry property (RIP), which is one of the most important tools in compressive sensing [33].

Notation: In this paper, bold-faced lower-case and upper-case letters stand for vectors and matrices, respectively. Superscripts  $\cdot^{\mathrm{T}}$ ,  $\cdot^{\mathrm{H}}$ , and  $\cdot^*$  denote transpose, Hermitian transpose, and complex conjugate operators, respectively. vec $\{\cdot\}$  denotes the vectorization operator, diag $\{\cdot\}$  returns a diagonal matrix whose main diagonal is given in the curly bracket, and Re $\{\cdot\}$  and Im $\{\cdot\}$  are real and imaginary parts of a complex-valued variable, respectively.  $\odot$  symbolizes the Khatri-Rao product.  $\mathbb{C}$  and  $\mathbb{R}$  are the sets of complex and real numbers, respectively. I is the identity matrix of appropriate dimension. 0 and 1 denote the all-zeros and the all-ones vectors of appropriate length, respectively. For a vector  $\mathbf{x}$ ,  $|\mathbf{x}|$  and  $||\mathbf{x}||_2$  represent the element-wise absolute value and the  $L_2$  norm of  $\mathbf{x}$ , respectively. The symbols  $\preceq$ ,  $\succeq$ , and  $\succ$  are element-wise less than or equal to, greater than or equal to, and greater than operators, respectively.

#### 2. Signal Model

Suppose a linear array of M sensors whose positions are contained in  $\mathbf{q} = [q_1, q_2, \cdots, q_M]^T$ , receives K far-field narrowband signals from directions  $\boldsymbol{\phi} = [\phi_1, \phi_2, \cdots, \phi_K]^T$  with  $\phi_k \in [-\pi/2, \pi/2)$ . For simplicity, we define the spatial frequencies as  $\boldsymbol{u} = [u_1, u_2, \cdots, u_K]^T$  with  $u_k = \sin(\phi_k) \in [-1, 1)$ . The array observation can be modeled as

$$\mathbf{y} = \sum_{k=1}^{K} s_k \mathbf{a}(u_k) + \mathbf{n} = \mathbf{A}(\boldsymbol{u})\mathbf{s} + \mathbf{n},$$

where  $s_k$  is the k-th signal waveform,  $\mathbf{s} = [s_1, s_2, \cdots, s_K]^T$  represents the signal vector, and  $\mathbf{n} \in \mathbb{C}^M$  is the noise vector. The steering matrix  $\mathbf{A}(\mathbf{u}) = [\mathbf{a}(u_1), \mathbf{a}(u_2), \cdots, \mathbf{a}(u_K)] \in \mathbb{C}^{M \times K}$  has the steering vectors as columns, where  $\mathbf{a}(u_k) = [e^{j\frac{2\pi q_1}{\lambda}u_k}, e^{j\frac{2\pi q_2}{\lambda}u_k}, \cdots, e^{j\frac{2\pi q_M}{\lambda}u_k}]^T$ , for  $k = 1, 2, \cdots, K$ , with  $\lambda$  being the signal wavelength and  $j = \sqrt{-1}$ . In grid-based methods, we formulate the signal model by means of a sparse representation, as

$$\mathbf{y} = \sum_{l=1}^{L} x_l \mathbf{a}(v_l) + \mathbf{n} = \mathbf{A}(\boldsymbol{v})\mathbf{x} + \mathbf{n},$$

where  $\boldsymbol{v} = [v_1, v_2, \cdots, v_L]^T$  denotes the frequency grid vector with L being the number of grids (in general  $L \gg M > K$ ),  $\mathbf{A}(\boldsymbol{v}) \in \mathbb{C}^{M \times L}$  stands for the overcomplete dictionary matrix, and  $\mathbf{x} = [x_1, x_2, \cdots, x_L]^T$  is a sparse vector whose elements  $x_l = s_k$  if  $v_l = u_k$ , and  $x_l = 0$  otherwise. When the true frequencies do not exactly lie in the preset grids, we encounter the off-grid issue. To handle this problem, we propose a method to simultaneously estimate the closest frequency grids, and the gaps between the closest grids and the true frequencies, using a second-order Taylor approximation.

# 3. Proposed Method

# 3.1. Second-Order Taylor Approximation

We start by considering a second-order Taylor approximation of the steering vectors. For any  $u_l$ , we have

$$\mathbf{a}(u_l) \approx \mathbf{a}(v_l) + \mathbf{a}'(v_l)p_l + \frac{\mathbf{a}''(v_l)}{2}p_l^2$$

where  $v_l$  is the grid closest to  $u_l$ ,  $\mathbf{a}'(v_l) = \left. \frac{\mathrm{d}\mathbf{a}(v)}{\mathrm{d}v} \right|_{v=v_l}$ ,  $\mathbf{a}''(v_l) = \left. \frac{\mathrm{d}^2\mathbf{a}(v)}{\mathrm{d}v^2} \right|_{v=v_l}$ , and  $p_l = u_l - v_l \in [-\delta/2, \delta/2]$  with  $\delta$  being the grid size. Collecting all the candidates, we have

$$[\mathbf{a}(u_1), \cdots, \mathbf{a}(u_L)] \approx \mathbf{A}(\mathbf{v}) + \mathbf{A}'(\mathbf{v}) \operatorname{diag}\{\mathbf{p}\} + \frac{1}{2}\mathbf{A}''(\mathbf{v}) \operatorname{diag}\{\mathbf{p}\}^2,$$

where  $\mathbf{A}'(\boldsymbol{v}) = [\mathbf{a}'(v_1), \cdots, \mathbf{a}'(v_L)] \in \mathbb{C}^{M \times L}, \mathbf{A}''(\boldsymbol{v}) = [\mathbf{a}''(v_1), \cdots, \mathbf{a}''(v_L)] \in \mathbb{C}^{M \times L}$ , and  $\mathbf{p} = [p_1, p_2, \cdots, p_L]^{\mathrm{T}}$ . Hence, the signal model can be approximately written as:

$$\mathbf{y} \approx \left[ \mathbf{A}(\boldsymbol{v}) + \mathbf{A}'(\boldsymbol{v}) \operatorname{diag}\{\mathbf{p}\} + \frac{1}{2} \mathbf{A}''(\boldsymbol{v}) \operatorname{diag}\{\mathbf{p}\}^2 \right] \mathbf{x} + \mathbf{n}$$
$$= \left[ \mathbf{A}(\boldsymbol{v}), \mathbf{A}'(\boldsymbol{v}), \frac{1}{2} \mathbf{A}''(\boldsymbol{v}) \right] \begin{bmatrix} \mathbf{x} \\ \operatorname{diag}\{\mathbf{p}\}\mathbf{x} \\ \operatorname{diag}\{\mathbf{p}\}^2 \mathbf{x} \end{bmatrix} + \mathbf{n},$$
(1)

where the signals of interest  $[\mathbf{x}^{\mathrm{T}}, (\operatorname{diag}\{\mathbf{p}\}\mathbf{x})^{\mathrm{T}}, (\operatorname{diag}\{\mathbf{p}\}^{2}\mathbf{x})^{\mathrm{T}}]^{\mathrm{T}}$  are referred to as block signal in the sequel.

# 3.2. Properties of the Block Signal

As shown in signal model (1), the unknown block signal is divided into three parts: (i)  $\mathbf{x}_1 \triangleq \mathbf{x}$ , (ii)  $\mathbf{x}_2 \triangleq \operatorname{diag}\{\mathbf{p}\}\mathbf{x}$ , and (iii)  $\mathbf{x}_3 \triangleq \operatorname{diag}\{\mathbf{p}\}^2\mathbf{x}$ . Without loss of generality, we assume  $\mathbf{x}$  is a real-valued vector, see Remark 1 below. Denote the *l*-th entries of  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ , and  $\mathbf{x}_3$  as  $x_{1,l}$ ,  $x_{2,l}$ , and  $x_{3,l}$ , respectively. We notice the following properties of the block signal  $[\mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}, \mathbf{x}_3^{\mathrm{T}}]^{\mathrm{T}}$ .

• Since  $\mathbf{x}$  is a sparse vector as mentioned in Section 2,  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ , and  $\mathbf{x}_3$  are all sparse and share the same sparsity pattern. This property is known as block-sparsity [13] or joint-sparsity [15].

• It holds that  $x_{2,l} = p_l x_{1,l}$  and  $x_{3,l} = p_l^2 x_{1,l}$ . Due to  $-\delta/2 \le p_l \le \delta/2$ ,  $\forall l \in \{1, 2, \dots, L\}$ , it is easy to verify that the following inequalities hold:

$$-\frac{\delta}{2}|\mathbf{x}_1| \leq \mathbf{x}_2 \leq \frac{\delta}{2}|\mathbf{x}_1|, \quad -\left(\frac{\delta}{2}\right)^2|\mathbf{x}_1| \leq \mathbf{x}_3 \leq \left(\frac{\delta}{2}\right)^2|\mathbf{x}_1|.$$
(2)

• It can be seen that  $\mathbf{x}_1, \, \mathbf{x}_2, \, \mathrm{and} \, \mathbf{x}_3$  satisfy the proportionality relationship, as

$$x_{2,l}^2 = x_{1,l} x_{3,l}, \ \forall l \in \{1, 2, \cdots, L\}.$$
(3)

**Remark 1.** For any complex-valued data model, say  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , we have its real-valued counterpart as  $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}$ , where  $\tilde{\mathbf{y}} = [\operatorname{Re}\{\mathbf{y}\}^{\mathrm{T}}, \operatorname{Im}\{\mathbf{y}\}^{\mathrm{T}}]^{\mathrm{T}}$ ,  $\tilde{\mathbf{x}} = [\operatorname{Re}\{\mathbf{x}\}^{\mathrm{T}}, \operatorname{Im}\{\mathbf{x}\}^{\mathrm{T}}]^{\mathrm{T}}$ , and  $\tilde{\mathbf{A}} = \begin{bmatrix} \operatorname{Re}\{\mathbf{A}\} & -\operatorname{Im}\{\mathbf{A}\} \\ \operatorname{Im}\{\mathbf{A}\} & \operatorname{Re}\{\mathbf{A}\} \end{bmatrix}$ .

# 3.3. Problem Formulation Development

Based on the aforementioned relationships among  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ , and  $\mathbf{x}_3$ , we propose the following minimization problem:

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3} \quad g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \quad \text{s.t.} \quad (2) \text{ and } (3). \tag{4}$$

The cost function in (4) is given by

$$g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \triangleq \frac{1}{2} \left\| \mathbf{y} - \mathbf{A}(\boldsymbol{v}) \mathbf{x}_1 - \mathbf{A}'(\boldsymbol{v}) \mathbf{x}_2 - \frac{1}{2} \mathbf{A}''(\boldsymbol{v}) \mathbf{x}_3 \right\|_2^2 + \mu \left\| \left[ \mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}, \mathbf{x}_3^{\mathrm{T}} \right]^{\mathrm{T}} \right\|_{2,1},$$
(5)

where  $\mu$  is a regularization parameter balancing the data fitting and the model sparsity, and  $\|\cdot\|_{2,1}$  is the mixed  $L_{2,1}$  norm of a vector, defined as

$$\left\| \left[ \mathbf{x}_{1}^{\mathrm{T}}, \mathbf{x}_{2}^{\mathrm{T}}, \mathbf{x}_{3}^{\mathrm{T}} \right]^{\mathrm{T}} \right\|_{2,1} = \sum_{l=1}^{L} \sqrt{|x_{1,l}|^{2} + |x_{2,l}|^{2} + |x_{3,l}|^{2}}.$$

Problem (4) is non-convex and hard to solve due to its constraints. We first consider the constraints of (2). The difficulty of dealing with (2) comes from the absolute value operator [15]. However, when the signals are assumed to be real positive, i.e.,  $\mathbf{s} \succ \mathbf{0}$  (and  $\mathbf{x}_1 = \mathbf{x} \succeq \mathbf{0}$ ), the constraints of (2) in (4) become

$$-\frac{\delta}{2}\mathbf{x}_1 \preceq \mathbf{x}_2 \preceq \frac{\delta}{2}\mathbf{x}_1, \quad \mathbf{0} \preceq \mathbf{x}_3 \preceq \left(\frac{\delta}{2}\right)^2 \mathbf{x}_1, \quad \mathbf{x}_1 \succeq \mathbf{0}, \tag{6}$$

which are linear and thus convex. It is worth pointing out that  $\mathbf{0} \leq \mathbf{x}_3$  in (6) is the result of  $x_{3,l} = p_l^2 x_{1,l}$ ,  $\forall l \in \{1, 2, \dots, L\}$  and  $\mathbf{x}_1 \succeq \mathbf{0}$ . Note that the assumption of real positive signals is valid in various situations. For instance, in multiple-snapshot scenarios, the signal vector denotes the signal powers which are naturally positive, see Remark 2.

In the sequel, we consider the last constraint in (4), viz. (3). Firstly, we convert (3) to its equivalent form as in [34]:

$$\left\| \begin{bmatrix} 2x_{2,l} \\ x_{1,l} - x_{3,l} \end{bmatrix} \right\|_{2} = x_{1,l} + x_{3,l}, \ \forall l \in \{1, 2, \cdots, L\}.$$
(7)

Then, we introduce an additional variable  $\mathbf{z} \in \mathbb{R}^L$  with entries  $z_l$  satisfying

$$0 \le z_l \le \eta, \ \forall l \in \{1, 2, \cdots, L\},\tag{8}$$

where  $\eta$  is a small user-defined parameter, and rewrite (7) as

$$\left\| \begin{bmatrix} 2x_{2,l} \\ x_{1,l} - x_{3,l} \end{bmatrix} \right\|_{2} \le x_{1,l} + x_{3,l} + z_{l}, \ \forall l \in \{1, 2, \cdots, L\},$$
(9)

which belongs to the set of standard second-order cone and hence is convex.

By replacing the constraint (2) with (6) and replacing (3) with (8) and (9), we finally relax the non-convex problem (4) into a convex one, as

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{z}} g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \qquad \text{s.t.} \ (6), \ (8), \ \text{and} \ (9).$$
(10)

**Remark 2.** Note that the proposed method is developed for the single-snapshot scenario. However, it can be easily extended to the case of multiple snapshots. To be precise, when multiple snapshots are available, we have the covariance matrix  $\mathbf{R} = \mathbf{A}\mathbf{R}_s\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}$ , where  $\sigma^2$  is the noise power. Note that we assume the signals to be uncorrelated with the noise, and the noise components are independent and identically distributed. Vectoring  $\mathbf{R}$  yields

$$\operatorname{vec}\{\mathbf{R}\} = (\mathbf{A}^* \odot \mathbf{A})\mathbf{r}_s + \sigma^2 \operatorname{vec}\{\mathbf{I}\},\tag{11}$$

where  $\mathbf{r}_s$  is the main diagonal of  $\mathbf{R}_s$ , denoting the signal powers. The data model (11) is similar to the signal model introduced in Section 2, and therefore, we can develop our method on the basis of (11).

To analyze the computational cost, we formulate Problem (10) under the framework of standard secondorder cone programming (SOCP) [35], as

$$\min_{\mathbf{x}_{1}, \mathbf{x}_{2}, \mathbf{x}_{3}, \mathbf{z}, \mathbf{t}} \sum_{l=1}^{L} t_{l} = \mathbf{1}^{\mathrm{T}} \mathbf{t}$$
s.t. (6), (8), and (9),  

$$\sqrt{|x_{1,l}|^{2} + |x_{2,l}|^{2} + |x_{3,l}|^{2}} \le t_{l}, \forall l \in \{1, 2, \cdots, L\},$$

$$\left\| \mathbf{y} - \mathbf{A}(\boldsymbol{v}) \mathbf{x}_{1} - \mathbf{A}'(\boldsymbol{v}) \mathbf{x}_{2} - \frac{1}{2} \mathbf{A}''(\boldsymbol{v}) \mathbf{x}_{3} \right\|_{2} \le \epsilon,$$

where  $\mathbf{t} = [t_1, t_2, \cdots, t_L]^{\mathrm{T}}$  is an auxiliary variable vector, and  $\epsilon$  is a tuning parameter related to  $\mu$  in (10). The computational cost of the above problem with implementation of SOCP is  $\mathcal{O}(9(M+1)L^2 + 72L)$  per iteration, and the number of iterations is bounded above by  $\mathcal{O}(\sqrt{L})$  [35]. The proposed second-order Taylor approximation method is referred to as 2nd Taylor G-LASSO. The computational complexity of the 2nd Taylor G-LASSO, as well as those of LASSO [1, 2], Neighbor G-LASSO [19], and 1st Taylor G-LASSO [13, 15], are summarized in Table 1.

Method	Cost per Iteration	No. of Iterations
LASSO	$\mathcal{O}((M+1)L^2)$	$\mathcal{O}(1)$
Neighbor G-LASSO	$\mathcal{O}\big(4(M+1)L^2+12L\big)$	$\mathcal{O}(\sqrt{L})$
1st Taylor G-LASSO	$\mathcal{O}\big(4(M+1)L^2+28L\big)$	$\mathcal{O}(\sqrt{L})$
2nd Taylor G-LASSO	$\mathcal{O}\big(9(M+1)L^2 + 72L\big)$	$\mathcal{O}(\sqrt{L})$

Table 1: Computational Cost Using SOCP Implementation

# 4. Uniqueness Property of the Proposed Solution

Note that, for underdetermined linear systems, uniqueness of a sparse solution is one of the fundamental problems in compressive sensing [36]. In this section, we discuss this issue in view of the proposed signal model in (1). To this end, we first introduce the following definition and theorem [33]:

**Definition.** An  $M \times bL$  block matrix **D** is said to have the block RIP with parameter  $\beta_K$ , if for every K block-sparse vector **c** of length bL, it holds that

$$(1 - \beta_K) \|\mathbf{c}\|_2^2 \le \|\mathbf{D}\mathbf{c}\|_2^2 \le (1 + \beta_K) \|\mathbf{c}\|_2^2$$

**Theorem.** Let  $\mathbf{y} = \mathbf{D}\mathbf{c}_0$  be measurements of a K block-sparse vector  $\mathbf{c}_0$ . If  $\mathbf{D}$  satisfies the block RIP with parameter  $\beta_{2K} < 1$ , then there exists a unique block-sparse vector  $\mathbf{c}$  satisfying  $\mathbf{y} = \mathbf{D}\mathbf{c}$ ; and further, if  $\mathbf{D}$  satisfies the block RIP with  $\beta_{2K} < \sqrt{2} - 1$ , then the convex optimization problem:  $\min_{\mathbf{c}} \|\mathbf{c}\|_{2,1}$  s.t.  $\mathbf{y} = \mathbf{D}\mathbf{c}$ , has a unique solution and the solution is equal to  $\mathbf{c}_0$ .

Define  $\mathbf{c}_0 = [\mathbf{x}^{\mathrm{T}}, (\operatorname{diag}\{\mathbf{p}\}\mathbf{x})^{\mathrm{T}}, (\operatorname{diag}\{\mathbf{p}\}^2\mathbf{x})^{\mathrm{T}}]^{\mathrm{T}}$  and  $\mathbf{D} = [\mathbf{A}(\boldsymbol{v}), \mathbf{A}'(\boldsymbol{v}), \frac{1}{2}\mathbf{A}''(\boldsymbol{v})]$ . In the absence of noise, our proposed model in (1) can be rewritten as:  $\mathbf{y} = \mathbf{D}\mathbf{c}_0$ . Without loss of generality, we denote  $\mathbf{\bar{D}}$  as the column-normalized matrix structured from  $\mathbf{D}$ . Our task is to check whether or not  $\mathbf{\bar{D}}$  satisfies the block RIP with parameter  $\beta_{2K} < 1$  and  $\beta_{2K} < \sqrt{2} - 1$ . Note that determining the RIP parameter, i.e.,  $\beta_{2K}$ , of a given matrix is in general an NP-hard problem [37, 38]. In what follows, we introduce a Monte Carlo test to check the condition of the block RIP of  $\mathbf{\bar{D}}$ .

According to the definition, if  $\overline{\mathbf{D}}$  has the block RIP with parameter  $\beta_{2K}$ , then for any 2K block-sparse vector  $\mathbf{c}$  of length bL, it holds that

$$(1 - \beta_{2K}) \|\mathbf{c}\|_{2}^{2} \leq \|\bar{\mathbf{D}}\mathbf{c}\|_{2}^{2} \leq (1 + \beta_{2K}) \|\mathbf{c}\|_{2}^{2}.$$
(12)

Note that, for any 2K block-sparse vector  $\mathbf{c}$ , we can write its unit-norm vector as  $\mathbf{\bar{c}} = \mathbf{c}/\|\mathbf{c}\|_2$ , such that  $\|\mathbf{\bar{c}}\|_2 = 1$ . As a result, (12) becomes:

$$(1 - \beta_{2K}) \le \frac{\|\bar{\mathbf{D}}\mathbf{c}\|_2^2}{\|\mathbf{c}\|_2^2} = \|\bar{\mathbf{D}}\bar{\mathbf{c}}\|_2^2 \le (1 + \beta_{2K}).$$

Based on the above inequalities, the parameter  $\beta_{2K}$  is calculated as

$$\beta_{2K} = \max\left\{ \|\bar{\mathbf{D}}\bar{\mathbf{c}}\|_2^2 - 1, 1 - \|\bar{\mathbf{D}}\bar{\mathbf{c}}\|_2^2 \right\}.$$
(13)

We randomly generate a unit-norm 2K block-sparse vector  $\bar{\mathbf{c}}$ , and calculate  $\beta_{2K}$  using (13). By repeatedly performing the above steps for  $10^4$  Monte Carlo runs, we estimate the empirical probabilities of  $\{\beta_{2K} < 1\}$ and  $\{\beta_{2K} < \sqrt{2} - 1\}$ . The empirical probabilities versus block-sparsity 2K are presented in Figure 1, with  $M = 8, q_m = \frac{(m-1)\lambda}{2}$  ( $m = 1, 2, \dots, M$ ), L = 200, and b = 1 for LASSO, b = 2 for Neighbor G-LASSO and 1st Taylor G-LASSO, and b = 3 for 2nd Taylor G-LASSO. It is seen that when the block-sparsity is small (less than 8), the probabilities of  $\{\beta_{2K} < 1\}$  of all the tested methods are high (greater than 0.9), and their probabilities of  $\{\beta_{2K} < \sqrt{2} - 1\}$  are larger than 0.5. Note that in Figure 1, the plot of 2nd Taylor G-LASSO with proportional signals (abbreviated as "Prop. Sig." in the figure), i.e., (3), has the highest probability. This reveals that the proportionality relationship of the block signal contains useful information in the sense that it increases the probabilities of  $\{\beta_{2K} < 1\}$  and  $\{\beta_{2K} < \sqrt{2} - 1\}$ .

# 5. Simulation

We evaluate the frequency estimation performance of 2nd Taylor G-LASSO, compared with LASSO [1–3], Neighbor G-LASSO [19], and 1st Taylor G-LASSO [13, 15]. We adopt the root-mean squared error (RMSE) and the empirical probability of correct detection (PCD) as performance metrics, defined as in [39]:

RMSE = 
$$10 \log_{10} \left( \sqrt{\frac{1}{KQ} \sum_{k=1}^{K} \sum_{q=1}^{Q} (\widehat{u}_{k,q} - u_k)^2} \right)$$

and PCD =  $Q_{\text{suc}}/Q$ , respectively, where  $\hat{u}_{k,q}$  denotes the frequency estimates of the k-th signal in the q-th Monte Carlo run, Q is the total number of Monte Carlo trials, and  $Q_{\text{suc}}$  is the number of trials where the frequency estimates  $\{\hat{u}_k | k = 1, 2, \dots, K\}$  fulfill:  $\max_k \{|\hat{u}_k - u_k|\} \leq \delta/2$ . The Cramér–Rao bound (CRB) [25] is drawn as a benchmark for RMSE comparison.

In the first experiment, a linear array of M = 16 omnidirectional sensors is considered to receive K = 2 signals with spatial frequencies  $u = [0.1815, 0.7942]^{\text{T}}$ . The M = 16 sensors are randomly selected from a ULA of 20 sensors with half-wavelength inter-element spacing. The frequency grid size is set to be  $\delta = 0.01$ , and hence the number of grids is L = 200. That is, the preset frequency grids are  $\{-1, -0.99, \dots, 0.98, 0.99\}$ . Two parameters utilized in (10) are given as  $\eta = 10^{-5}$  and  $\mu = \sigma \sqrt{M \ln(M)}$  [40] with  $\sigma$  denoting the standard deviation of the noise vector, which is assumed to be known a priori in our simulations. Q = 1000 Monte Carlo trials are performed. The results of RMSE versus SNR and PCD versus SNR are plotted in Figures 2 and 3, respectively. It is seen that, in the large SNR region, 2nd Taylor G-LASSO has significantly lower RMSE compared with the other grid-based approaches, and the PCD of 2nd Taylor G-LASSO is higher than those of the other tested methods.

In the second experiment, we randomly select M sensors from a ULA of 20 sensors with half-wavelength inter-element spacing, and M varies from 4 to 20. SNR is fixed to 20 dB, while the remaining parameters are the same as those in the first experiment. The RMSE and PCD are depicted in Figures 4 and 5, respectively. The results exhibit again better performance of the proposed 2nd Taylor G-LASSO than the other competitors. In the third experiment, the number of frequency grids, i.e., L, varies from 50 to 500 with a step size of 50, the SNR is fixed to 20 dB, while the other parameters are unchanged as those in the first experiment. The RMSE and PCD results are shown in Figures 6 and 7, respectively. It can be seen that (i) When the number of grids is L < 400 (equivalently grid size of  $\delta > 1/200$ ), the RMSE of 2nd Taylor G-LASSO is evidently smaller than those of the other tested methods; and (ii) When  $L \ge 400$  (that is  $\delta \le 1/200$ ), the RMSE of 1st Taylor G-LASSO is very close to that of 2nd Taylor G-LASSO. This verifies that 2nd Taylor G-LASSO works better than 1st Taylor G-LASSO in terms of DOA estimation accuracy, especially when the grid size is not sufficiently small.

In the last experiment, we test the performances of the proposed method and several other algorithms in multiple-snapshot scenarios. We utilize 100 snapshots, and the other parameters are set to be the same as those in the first experiment. The strategy of transforming the multiple-snapshot signal model into a single-snapshot one, which has been detailed in Remark 2 in Section 3.3, is applied to LASSO, Neighbor G-LASSO, 1st Taylor G-LASSO, and 2nd Taylor G-LASSO. In addition, in this example, we also consider two classical methods, namely, the Capon beamforming and multiple signal classification (MUSIC) algorithms [41]. For comparison, on-grid MUSIC with a much tinier grid size  $\delta = 0.0001$  is also examined. The RMSE and PCD are plotted in Figures 8 and 9, respectively, from which it is seen that both Capon beamforming and MUSIC algorithms share similar performance with LASSO in the off-grid setup. On-grid MUSIC has the smallest RMSE and the largest PCD among all the tested approaches. The proposed 2nd Taylor G-LASSO outperforms LASSO, Neighbor G-LASSO, and 1st Taylor G-LASSO.

# 6. Conclusion

We have investigated the off-grid DOA estimation problem and have proposed a method using the secondorder Taylor approximation. By exploring the properties of the block signal, we have added the proportionality relationship to our optimization problem. A Monte Carlo test has shown the usefulness of such proportionality relationship in the sense that it increases the probabilities of  $\{\beta_{2K} < 1\}$  and  $\{\beta_{2K} < \sqrt{2}-1\}$ . Numerical results have demonstrated that the proposed method outperforms several existing grid-based DOA estimation approaches.

# References

- R. Tibshirani, Regression shrinkage and selection via the LASSO, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288. doi:https://www.jstor.org/stable/2346178.
- [2] D. Malioutov, M. Çetin, A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Transactions on Signal Processing 53 (8) (2005) 3010–3022. doi:10.1109/TSP.2005.850882.

- [3] C. Weiss, A. M. Zoubir, Robust high-resolution DOA estimation with array pre-calibration, in: Proceedings of European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 2014, pp. 1049–1052.
- [4] P. Stoica, P. Babu, J. Li, New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data, IEEE Transactions on Signal Processing 59 (1) (2011) 35-47. doi:10.1109/TSP.2010.2086452.
- [5] P. Stoica, P. Babu, J. Li, SPICE: A sparse covariance-based estimation method for array processing, IEEE Transactions on Signal Processing 59 (2) (2011) 629–638. doi:10.1109/TSP.2010.2090525.
- [6] P. Stoica, P. Babu, SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation, Signal Processing 92 (7) (2012) 1580–1590. doi:10.1016/j.sigpro.2011.11.010.
- P. Babu, P. Stoica, Connection between SPICE and square-root LASSO for sparse parameter estimation, Signal Processing 95 (2014) 10–14. doi:10.1016/j.sigpro.2013.08.011.
- [8] Z. Yang, J. Li, P. Stoica, L. Xie, Chapter 11 Sparse methods for direction-of-arrival estimation, in: R. Chellappa, S. Theodoridis (Eds.), Academic Press Library in Signal Processing, Volume 7, Academic Press, 2018, pp. 509–581. doi:https://doi.org/10.1016/B978-0-12-811887-0.00011-0.
- [9] H. Zhu, G. Leus, G. B. Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling, IEEE Transactions on Signal Processing 59 (5) (2011) 2002–2016.
- [10] Z. Yang, C. Zhang, L. Xie, Robustly stable signal recovery in compressed sensing with structured matrix perturbation, IEEE Transactions on Signal Processing 60 (9) (2012) 4658–4671.
- [11] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, IEEE Transactions on Signal Processing 61 (1) (2013) 38–43. doi:10.1109/TSP.2012.2222378.
- [12] M. F. Duarte, R. G. Baraniuk, Spectral compressive sensing, Applied and Computational Harmonic Analysis 35 (1) (2013) 111–129. doi:10.1016/j.acha.2012.08.003.
- [13] R. Jagannath, K. Hari, Block sparse estimator for grid matching in single snapshot DoA estimation, IEEE Signal Processing Letters 20 (11) (2013) 1038–1041. doi:10.1109/LSP.2013.2279124.
- [14] S. Fortunati, R. Grasso, F. Gini, M. S. Greco, K. LePage, Single-snapshot DOA estimation by using compressed sensing, EURASIP Journal on Advances in Signal Processing 120 (2014) 1–17. doi:10.1186/1687-6180-2014-120.
- [15] Z. Tan, P. Yang, A. Nehorai, Joint sparse recovery method for compressed sensing with structured dictionary mismatches, IEEE Transactions on Signal Processing 62 (19) (2014) 4997–5008. doi:10.1109/TSP.2014.2343940.

- [16] J. Dai, X. Bao, W. Xu, C. Chang, Root sparse Bayesian learning for off-grid DOA estimation, IEEE Signal Processing Letters 24 (1) (2017) 46–50. doi:10.1109/LSP.2016.2636319.
- [17] A. C. Walewski, C. Steffens, M. Pesavento, Off-grid parameter estimation based on joint sparse regularization, in: Proceedings of International ITG Conference on Systems, Communications and Coding (SCC), Hamburg, Germany, 2017, pp. 1–6.
- [18] Q. Liu, H. C. So, Y. Gu, Off-grid DOA estimation with nonconvex regularization via joint sparse representation, Signal Processing 140 (2017) 171–176. doi:10.1016/j.sigpro.2017.05.020.
- [19] A. Abtahi, S. Gazor, F. Marvasti, Off-grid localization in MIMO radars using sparsity, IEEE Signal Processing Letters 25 (2) (2018) 313–317. doi:10.1109/LSP.2018.2791447.
- [20] Q. Wang, Z. Zhao, Z. Chen, Z. Nie, Grid evolution method for DOA estimation, IEEE Transactions on Signal Processing 66 (9) (2018) 2374–2383. doi:10.1109/TSP.2018.2814998.
- [21] X. Wu, W.-P. Zhu, J. Yan, Z. Zhang, Two sparse-based methods for off-grid direction-of-arrival estimation, Signal Processing 142 (2018) 87–95. doi:10.1016/j.sigpro.2017.07.004.
- [22] C. Zhou, Y. Gu, Z. Shi, Y. D. Zhang, Off-grid direction-of-arrival estimation using coprime array interpolation, IEEE Signal Processing Letters 25 (11) (2018) 1710–1714.
- [23] B. Wang, Y. Gu, W. Wang, Off-grid direction-of-arrival estimation based on steering vector approximation, Circuits, Systems, and Signal Processing 38 (3) (2019) 1287–1300. doi:10.1007/s00034-018-0914-5.
- [24] Z. Wan, W. Liu, Non-coherent DOA estimation of off-grid signals with uniform circular arrays, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Canada, 2021, pp. 4370–4374. doi:10.1109/ICASSP39728.2021.9415069.
- [25] Y. Ma, X. Cao, X. Wang, M. S. Greco, F. Gini, Multi-source off-grid DOA estimation with single snapshot using non-uniform linear arrays, Signal Processing 189 (2021) 108238.
- [26] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Transactions on Information Theory 59 (11) (2013) 7465–7490. doi:10.1109/TIT.2013.2277451.
- [27] Z. Yang, L. Xie, On gridless sparse methods for line spectral estimation from complete and incomplete data, IEEE Transactions on Signal Processing 63 (12) (2015) 3139–3153. doi:10.1109/TSP.2015.2420541.
- [28] J. Steinwandt, F. Roemer, C. Steffens, M. Haardt, M. Pesavento, Gridless super-resolution direction finding for strictly non-circular sources based on atomic norm minimization, in: Proceedings of 50th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, USA, 2016, pp. 1518–1522. doi:10.1109/ACSSC.2016.7869631.

- [29] C. Steffens, W. Suleiman, A. Sorg, M. Pesavento, Gridless compressed sensing under shift-invariant sampling, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, 2017, pp. 4735–4739. doi:10.1109/ICASSP.2017.7953055.
- [30] M. Wagner, P. Gerstoft, Y. Park, Gridless DOA estimation via. alternating projections, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 4215–4219. doi:10.1109/ICASSP.2019.8682920.
- [31] J. Zhang, D. Rakhimov, M. Haardt, Gridless channel estimation for hybrid mmWave MIMO systems via tensor-ESPRIT algorithms in DFT beamspace, IEEE Journal of Selected Topics in Signal Processing 15 (3) (2021) 816-831. doi:10.1109/JSTSP.2021.3063908.
- [32] M. Wagner, Y. Park, P. Gerstoft, Gridless DOA estimation and root-MUSIC for non-uniform linear arrays, IEEE Transactions on Signal Processing 69 (2021) 2144–2157. doi:10.1109/TSP.2021.3068353.
- [33] Y. C. Eldar, M. Mishali, Robust recovery of signals from a structured union of subspaces, IEEE Transactions on Information Theory 55 (11) (2009) 5302–5316. doi:10.1109/TIT.2009.2030471.
- [34] J. Park, S. Boyd, General heuristics for nonconvex quadratically constrained quadratic programming, arXiv: Optimization and Control (2017). arXiv:1703.07870.
- [35] M. S. Lobo, L. Vandenberghe, S. Boyd, H. Lebret, Applications of second-order cone programming, Linear Algebra and Its Applications 284 (1) (1998) 193–228. doi:10.1016/S0024-3795(98)10032-0.
- [36] S. Foucart, H. Rauhut, Chapter 2 Sparse solutions of underdetermined systems, in: A Mathematical Introduction to Compressive Sensing, Springer New York, 2013, pp. 41–59. doi:10.1007/978-0-8176-4948-7\_2.
- [37] Y. C. Eldar, P. Kuppinger, H. Bölcskei, Block-sparse signals: Uncertainty relations and efficient recovery, IEEE Transactions on Signal Processing 58 (6) (2010) 3042–3054. doi:10.1109/TSP.2010.2044837.
- [38] A. M. Tillmann, M. E. Pfetsch, The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing, IEEE Transactions on Information Theory 60 (2) (2014) 1248–1259. doi:10.1109/TIT.2013.2290112.
- [39] C. Steffens, M. Pesavento, Block- and rank-sparse recovery for direction finding in partly calibrated arrays, IEEE Transactions on Signal Processing 66 (2) (2018) 384–399. doi:10.1109/TSP.2017.2770104.
- [40] B. N. Bhaskar, G. Tang, B. Recht, Atomic norm denoising with applications to line spectral estimation, IEEE Transactions on Signal Processing 61 (23) (2013) 5987–5999. doi:10.1109/TSP.2013.2273443.
- [41] P.-J. Chung, M. Viberg, J. Yu, Chapter 14 DOA estimation methods and algorithms, in: A. M. Zoubir, M. Viberg, R. Chellappa, S. Theodoridis (Eds.), Academic Press Library in Signal Processing, Volume 3, Elsevier, 2014, pp. 599–650. doi:https://doi.org/10.1016/B978-0-12-411597-2.00014-X.



Figure 1: Empirical probabilities of  $\{\beta_{2K} < 1\}$  and  $\{\beta_{2K} < \sqrt{2} - 1\}$  versus block-sparsity 2K with 10<sup>4</sup> Monte Carlo runs, M = 8, and L = 200.



Figure 2: RMSE versus SNR with M = 16 sensors, K = 2 sources, L = 200 frequency grids, and grid size  $\delta = 0.01$ .



Figure 3: PCD versus SNR with M = 16 sensors, K = 2 sources, L = 200 frequency grids, and grid size  $\delta = 0.01$ .

![](_page_14_Figure_0.jpeg)

Figure 4: RMSE versus number of sensors with SNR = 20 dB, K = 2 sources, L = 200 frequency grids, and grid size  $\delta = 0.01$ .

![](_page_15_Figure_0.jpeg)

Figure 5: PCD versus number of sensors with SNR = 20 dB, K = 2 sources, L = 200 frequency grids, and grid size  $\delta = 0.01$ .

![](_page_16_Figure_0.jpeg)

Figure 6: RMSE versus number of frequency grids with SNR = 20 dB, M = 16 sensors, and K = 2 sources.

![](_page_17_Figure_0.jpeg)

Figure 7: PCD versus number of frequency grids with SNR = 20 dB, M = 16 sensors, and K = 2 sources.

![](_page_18_Figure_0.jpeg)

Figure 8: RMSE versus SNR in multiple-snapshot scenarios with 100 snapshots, M = 16 sensors, and K = 2 sources.

![](_page_19_Figure_0.jpeg)

Figure 9: PCD versus SNR in multiple-snapshot scenarios with 100 snapshots, M = 16 sensors, and K = 2 sources.