Simultaneous Signal Subspace Rank and Model Selection with an Application to Single-snapshot Source Localization

Muhammad Naveed Tabassum and Esa Ollila

Aalto University, Dept. of Signal Processing and Acoustics, P.O. Box 15400, FI-00076 Aalto, Finland

Email: {muhammad.tabassum, esa.ollila}@aalto.fi

Abstract—**This paper proposes a novel method for model selection in linear regression by utilizing the solution path of ℓ_1 regularized least-squares (LS) approach (i.e., Lasso). This method applies the complex-valued least angle regression and shrinkage (c-LARS) algorithm coupled with a generalized information criterion (GIC) and referred to as the c-LARS-GIC method. c-LARS-GIC is a two-stage procedure, where firstly precise values of the regularization parameter, called knots, at which a new predictor variable enters (or leaves) the active set are computed in the Lasso solution path. Active sets provide a nested sequence of regression models and GIC then selects the best model. The sparsity order of the chosen model serves as an estimate of the model order and the LS fit based only on the active set of the model provides an estimate of the regression parameter vector. We then consider a source localization problem, where the aim is to detect the number of impinging source waveforms at a sensor array as well to estimate their direction-of-arrivals (DoAs) using only a single-snapshot measurement. We illustrate via simulations that, after formulating the problem as a grid-based sparse signal reconstruction problem, the proposed c-LARS-GIC method detects the number of sources with high probability while at the same time it provides accurate estimates of source locations.**

I. INTRODUCTION

We assume that a measurement vector $\mathbf{y} \in \mathbb{C}^n$ is generated via linear model

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (1)$$

where $\mathbf{\Phi} = \begin{pmatrix} \boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_p \end{pmatrix} \in \mathbb{C}^{n \times p}$ is the known basis matrix (or matrix of predictors), $\boldsymbol{\beta} \in \mathbb{C}^p$ is the unknown signal vector (or regression coefficient vector) and $\boldsymbol{\varepsilon} \in \mathbb{C}^n$ is the (unobserved) random noise vector. We assume that the model is underdetermined, i.e., $p > n$, and the signal vector $\boldsymbol{\beta}$ is *sparse* (i.e., having only a few non-zero elements) with sparsity order $k^* = \|\boldsymbol{\beta}\|_0 \ll n$, where $\|\cdot\|_0$ denotes the ℓ_0-(pseudo)norm, $\|\boldsymbol{\beta}\|_0 = |\{i : \beta_i \neq 0\}|$, i.e., $\|\boldsymbol{\beta}\|_0$ is equal to number of non-zero elements of $\boldsymbol{\beta}$. Such a model arises in compressed sensing [1] and in sparse linear regression [2]. For ease of exposition, we consider centered linear model, i.e., the intercept is equal to zero.

In this paper, we consider solving the following tasks

- *Detection task*, where the problem is to find the true model (sparsity) order k^*, or in other words, to detect the rank of the signal subspace.
- *Sparse signal reconstruction (SSR) task*, where the problem is to estimate the unknown k^*-sparse signal $\boldsymbol{\beta}$.

So far most research has focused on SSR task and much less attention has been paid to solve the detection task. Some existing detection approaches in the literature can be found e.g., in [3], [4]. Typically, most SSR (e.g., greedy pursuit methods [5]) presume that k^* is known or an estimate of it is available. In practice, the underlying sparsity order k^* is typically unknown and can even vary with time, e.g., as new measurements (or snapshots) become available. Thus the performance of greedy pursuit methods depends heavily on the accuracy of the method used for solving the detection task. Some other approaches, such as sparse linear regression methods [2], perform detection and SSR task simultaneously. For example, the Least absolute shrinkage and selection operator (Lasso) [6] uses a data-dependent penalty parameter λ commonly chosen by cross-validation to compute an estimate $\hat{\boldsymbol{\beta}}$ of sparse vector $\boldsymbol{\beta}$. An estimate of the sparsity order k^* is then obtained as $\hat{k}^* = \|\hat{\boldsymbol{\beta}}\|_0$.

The Lasso estimator of $\boldsymbol{\beta}$ solves the following penalized (regularized) residual sum of squares (RSS) optimization problem,

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta} \in \mathbb{C}^p}{\arg\min} \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \qquad (2)$$

where $\lambda \geq 0$ is the penalty parameter that controls the sparsity order of the solution. The *least angle regression and shrinkage (LARS)* algorithm [7] is a novel method for computing the Lasso solution (regularization) path as λ varies. The LARS method finds the sequence of values λ_k, referred to as *knots*, at which a new predictor variable enters (or leaves) the active set of non-zero coefficients. Additionally, it provides the respective solutions $\hat{\boldsymbol{\beta}}(\lambda_k)$ for $k = 0, 1, \ldots, K = \min(n-1, p)$. Herein, we use our complex-valued extension of LARS (see [8], [9]) to find the knots and the associated active sets of non-zero coefficients in the Lasso solution path [7], [8], [10]. Active sets then provide a nested sequence of regression models. At the second stage, we then apply an information criterion (IC) to choose the best model among these candidate models. This approach is referred to as *c-LARS-GIC* method in the sequel.

There are several information criteria (e.g., Akaike or Bayesian or their extensions) proposed in the literature (see [11] for a review) that can be utilized at the second stage. Another alternative would be to use sequential generalized

likelihood ratio tests [12], the covariance test [13] or spacing test [14], for example. However, we choose to use information criteria due to their simplicity and ease of implementation. Furthermore, tests developed in [13], [14] work only for real-valued data.

In this paper, we use *generalized information criterion (GIC)* formulation [15] that contains both the Bayesian information criterion (BIC) [16], and its modifications developed in [17], [18] as special cases. The sparsity order of the chosen model serves as an estimate of the model order, and the least-squares (LS) fit based on the active set of the chosen model provides estimates of the non-zero signal coefficients. Thus, *c-LARS-GIC* approach solves both the detection and estimation task simultaneously.

This paper is organized as follows. Section II develops the c-LARS-GIC method. The grid-based SSR framework for single-snapshot source localization problem using a sensor array, referred to as compressed beamforming (CBF), is described in Section III. The results for a set of simulation studies then illustrate the effectiveness of the c-LARS-GIC approach in CBF application. Section IV discusses the key outcomes and concludes the paper.

Notations: Lowercase boldface letters are used for vectors and uppercase for matrices. The ℓ_2 and ℓ_1 norms of the vector $\mathbf{a} \in \mathbb{C}^p$ are defined as $\|\mathbf{a}\|_2 = \sqrt{\mathbf{a}^{\mathsf{H}}\mathbf{a}}$ and $\|\mathbf{a}\|_1 = \sum_{j=1}^{p}|a_i|$, where $|a| = \sqrt{a^*a} = \sqrt{a_R^2 + a_I^2}$ denotes the modulus of a complex number $a = a_R + \jmath a_I$ and $(\cdot)^{\mathsf{H}} = [(\cdot)^*]^{\top}$ denotes the Hermitian (complex conjugate) transpose. The support of $\mathbf{a}$ is the index set of its non-zero elements, i.e., $\mathcal{A} = \mathrm{supp}(\mathbf{a}) = \{j \in \{1, \dots, p\} : a_j \neq 0\}$. The ℓ_0-(pseudo)norm of $\mathbf{a}$ is defined as $\|\mathbf{a}\|_0 = |\mathrm{supp}(\mathbf{a})|$, which is equal to the total number of non-zero elements in it. If $\boldsymbol{\beta} \in \mathbb{C}^p$, $\boldsymbol{\Phi} \in \mathbb{C}^{n \times p}$ and $\mathcal{A}$ is an index set $\mathcal{A} \subset \{1, \dots, n\}$ of cardinality $|\mathcal{A}| = k$, then $\boldsymbol{\beta}_{\mathcal{A}} \in \mathbb{C}^k$ (resp. $\boldsymbol{\Phi}_{\mathcal{A}}$) denotes the sub-vector (resp. $n \times k$ sub-matrix) restricted to components of $\boldsymbol{\beta}$ (resp. columns of $\boldsymbol{\Phi}$) indexed by the set $\mathcal{A}$. The $\boldsymbol{\Phi}^+$ is a Moore-Penrose pseudoinverse of $\boldsymbol{\Phi}$. We use $\mathsf{I}(\cdot)$ to denote the indicator function and $\mathrm{ave} = \frac{1}{M}\sum_{m=1}^{M}(\cdot)$ for M numbers.

II. THE C-LARS-GIC METHOD

A. The Lasso Knots and Active Sets

Recall that the signal vector $\boldsymbol{\beta} \in \mathbb{C}^p$ is sparse, so only few elements are non-zero. The support $\mathcal{A}^*$ of $\boldsymbol{\beta}$ is the index set of its non-zero elements, i.e., $\mathcal{A}^* = \mathrm{supp}(\boldsymbol{\beta})$ and the cardinality of this set, $k^* = \|\boldsymbol{\beta}\|_0 < n < p$, is the true sparsity level or the model order. The main aim in detection task is to find the true order k^* whereas the main interest in SSR task is on accurate variable selection, i.e., on identifying the true active set $\mathcal{A}^*$. Naturally, if $\mathcal{A}^*$ would be known, then one can simply compute the LS solution when regressing $\mathbf{y}$ on $\boldsymbol{\Phi}_{\mathcal{A}^*}$, where $\boldsymbol{\Phi}_{\mathcal{A}^*}$ denotes the $n \times k^*$ (sub)matrix of $\boldsymbol{\Phi}$ restricted to the columns of the active set $\mathcal{A}^*$.

Recall that Lasso estimate $\hat{\boldsymbol{\beta}}(\lambda) \in \mathbb{C}^p$ is sparse having *at most n non-zero elements*. The sparsity level of $\hat{\boldsymbol{\beta}}(\lambda)$ depends on the penalty parameter λ. We assume that the columns of

$\boldsymbol{\Phi}$ are centered, i.e., $\|\boldsymbol{\phi}_j\|^2 = 1$ holds. In Lasso estimation framework this can be done without loss of generality since it only implies that the found Lasso solution (based on normalized basis vectors) is rescaled back to the original scale of $\boldsymbol{\phi}_j$-s; see [2] for more elaborate discussion of this feature of the Lasso.

Let λ_0 denotes the smallest value of λ such that all coefficients are zero, i.e., $\hat{\boldsymbol{\beta}}(\lambda_0) = \mathbf{0}_p$. It is easy to see that [2]

$$\lambda_0 = \max_{j \in \{1, \dots, p\}} |\langle \boldsymbol{\phi}_j, \mathbf{y} \rangle|. \tag{3}$$

Let $\mathcal{A}(\lambda) = \mathrm{supp}\{\hat{\boldsymbol{\beta}}(\lambda)\}$ denotes the *active set* at the regularization parameter value $\lambda < \lambda_0$. The *knots* $\lambda_0 > \lambda_1 > \lambda_2 > \cdots > \lambda_K$ are defined as the smallest values of the regularization parameter after which there is a change in the set of active predictors, i.e., the order of sparsity changes. The active set at a knot λ_k is denoted by $\mathcal{A}_k = \mathcal{A}(\lambda_k) = \mathrm{supp}\{\hat{\boldsymbol{\beta}}(\lambda_k)\}$. The active set $\mathcal{A}_1$ thus contains a single index as $\mathcal{A}_1 = \{j_1\}$, where j_1 is predictor that becomes active first, i.e.,

$$j_1 = \arg\max_{j \in \{1, \dots, p\}} |\langle \boldsymbol{\phi}_j, \mathbf{y} \rangle|.$$

By definition of the knots, one has that

$$\begin{aligned} \mathcal{A}_k &= \mathrm{supp}\{\hat{\boldsymbol{\beta}}(\lambda)\} && \forall\, \lambda_{k-1} < \lambda \le \lambda_k \\ \mathcal{A}_k &\neq \mathcal{A}_{k+1}, && \forall\, k = 0, 1, \dots, K \end{aligned}$$

and $\mathcal{A}_0 = \mathrm{supp}\{\hat{\boldsymbol{\beta}}(\lambda)\} = \{\emptyset\}$ for all $\lambda \ge \lambda_0$.

Let $\mathcal{A} = (i_1, \dots, i_k) \subset \{1, \dots, n\}$ denotes an index set of variables (basis vectors $\boldsymbol{\phi}_j$-s) that are included in the model and suppose that its sparsity order is $k = |\mathcal{A}| < n$. Then $\mathcal{A}^{\complement} = \{1, \dots, p\} \setminus \mathcal{A}$ is the index set of signal coefficients that are zero. Then, ideally, we would like to consider all hypotheses of the form

$$H_{\mathcal{A}} : \boldsymbol{\beta}_{\mathcal{A}^{\complement}} = \mathbf{0}, \qquad |\mathcal{A}| = k \implies |\mathcal{A}^{\complement}| = p - k \tag{4}$$

and choose the best model among all possible models $\{H_{\mathcal{A}}\}$ as the one that appears most plausible based on the data. The posed hypotheses testing problem is obviously computationally infeasible even for small n. Indeed, observe that there are $\binom{n}{k}$ index sets $\mathcal{A}$ of size $|\mathcal{A}| = k$, where $k = 1, \dots, n$. Thus it is not possible to go through all possible models since the total number of all models is $O(n^n)$.

Finding the Lasso knots $\lambda_0, \lambda_1, \dots, \lambda_K$ and the corresponding active sets $\mathcal{A}_k$ for $k = 0, 1, \dots, K$, via the c-LARS-WLasso algorithm of [8], [9] with unit weights allows us to form a nested set of hypotheses,

$$\begin{aligned} H_0 &\subset H_1 \subset \cdots \subset H_K \\ H_k &: \boldsymbol{\beta}_{\mathcal{A}_k^{\complement}} = \mathbf{0}, \qquad k = 0, 1, \dots, K. \end{aligned} \tag{5}$$

This reduces the number of tested hypotheses from n^n to $(K+1) \le n \le p$.

B. Model Selection using Information Criteria

Information criteria are a common way of choosing among models while balancing the competing goals of goodness-of-fit and parsimony (simpler model). The Generalized information criterion (GIC) [15] for hypothesis $H_{\mathcal{A}}$ is

$$\text{GIC}_\gamma(\mathcal{A}) = n \ln \hat{\sigma}^2(\mathcal{A}) + k \, c_{n,\gamma} \qquad (6)$$

where $k = |\mathcal{A}|$ is the sparsity (model) order of hypotheses $H_{\mathcal{A}}$, and

$$\hat{\sigma}^2(\mathcal{A}) = \frac{1}{n} \left\| \mathbf{y} - \boldsymbol{\Phi}_{\mathcal{A}} \hat{\mathbf{s}} \right\|_2^2, \quad \hat{\mathbf{s}} = \boldsymbol{\Phi}_{\mathcal{A}}^+ \mathbf{y}$$

are the ML-estimators of error scale $\sigma^2 > 0$ and $\boldsymbol{\beta}_{\mathcal{A}^*}$ under $H_{\mathcal{A}}$ in (4) and assuming that $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$. Furthermore, $c_{n,\gamma}$ in (6) is a positive *sequence* depending on n. Subscript γ obtains values $\gamma \in \{0, 1, 2\}$ and is used to indicate the different options of sequence $c_{n,\gamma}$ suggested in the literature:

- $\text{GIC}_0 = \text{BIC}$ [16] uses $c_{n,0} = \ln n$.
- GIC_1 uses $c_{n,1} = \ln n \cdot \ln(\ln p)$ as suggested in [17].
- GIC_2 uses $c_{n,2} = \ln p \cdot \ln(\ln n)$ as suggested in [18].

If we assume that the set of hypotheses $\{H_{\mathcal{A}}\}$ includes the true model $H_{\mathcal{A}^*}$ that generated the data, then the GIC estimates used above have proven to be consistent [15], [17], [18] under some technical conditions, that is, $\mathbb{P}(\text{``correctly choosing } H_{\mathcal{A}^*}\text{''}) \to 1$ as $n \to \infty$. Finally, we point out that the ML estimate $\hat{\sigma}^2 = \hat{\sigma}^2(\mathcal{A})$ of σ^2 under the hypothesis $H_{\mathcal{A}}$ is biased. Therefore, we instead use the bias corrected (i.e., unbiased) estimate, given as:

$$\hat{\sigma}_u^2(\mathcal{A}) = \frac{1}{n-k} \left\| \mathbf{y} - \boldsymbol{\Phi}_{\mathcal{A}} \hat{\mathbf{s}} \right\|_2^2, \qquad (7)$$
$$k = |\mathcal{A}|, \qquad \hat{\mathbf{s}} = \boldsymbol{\Phi}_{\mathcal{A}}^+ \mathbf{y}.$$

Note that we will use $\hat{\sigma}_u^2(\mathcal{A}_k)$ in place of $\hat{\sigma}^2(\mathcal{A})$ in (6), for nested hypothesis set H_k of (5), where $k = 0, 1, \ldots, K$.

C. The c-LARS-GIC Method

We are now ready to state the c-LARS-GIC procedure for testing the nested set of hypotheses in (5) for choosing the correct model and the sparsity order. The method is described in Algorithm 1.

It is important to note that the nested set of hypotheses (5) are not pre-fixed as usually but are based on the data. Naturally, the success of the c-LARS-GIC method depends on the ability of the Lasso to have the correct model $\mathcal{A}^*$ among the active sets $\{\mathcal{A}_k\}_{k=1}^K$ which are determined at the knots in the Lasso path. Nevertheless, when the true $\mathcal{A}^*$ is included in the set of models, then c-LARS-GIC has a high probability of choosing the correct model due to consistency property of the used information criterion. This feature is indeed verified in our simulation studies in Section III-B.

D. Conventional Approach

We compare the c-LARS-GIC method to conventional approach where Lasso estimates $\{\hat{\boldsymbol{\beta}}(\lambda_l)\}_{l=0}^L$ are found on a grid of λ values, e.g., by cyclic coordinate descent algorithm [2],

$$\{\lambda_0, \ldots, \lambda_L\}, \quad \lambda_0 > \lambda_1 > \ldots > \lambda_L,$$

Algorithm 1: The c-LARS-GIC method.

input	:	$\mathbf{y} \in \mathbb{C}^n$, $\boldsymbol{\Phi} \in \mathbb{C}^{n \times p}$ and $\gamma \in \{0, 1, 2\}$.
output	:	$(\hat{k}^*, \hat{\mathcal{A}}^*, \hat{\boldsymbol{\beta}})$.
initialize :		$\hat{\boldsymbol{\beta}} = \mathbf{0}_p$ and $K = \min(n-1, p)$.

1 Compute the knots λ_k and the corresponding active sets $\mathcal{A}_k$ using the c-LARS-WLasso algorithm with unit weights, where $k = 0, 1, \ldots, K$; see [8] and [9].

2 Find the sparsity level k^* using (6) as:

$$\hat{k}^* = \underset{k \in \{0,1,\ldots,K\}}{\arg\min} \{ \text{GIC}_\gamma(\lambda_k) = n \ln \hat{\sigma}_u^2(\mathcal{A}_k) + k \, c_{n,\gamma} \},$$

where $\hat{\sigma}_u^2(\mathcal{A}_k)$ is computed using (7). Thus, an estimate of the active set is $\hat{\mathcal{A}}^* = \mathcal{A}_{\hat{k}^*}$.

3 Set $\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}^*} = \boldsymbol{\Phi}_{\hat{\mathcal{A}}^*}^+ \mathbf{y}$ to get $\hat{k}^*$-sparse estimate $\hat{\boldsymbol{\beta}} \in \mathbb{C}^p$.

---

where the sequence $\{\lambda_l\}$ is monotonically decreasing from λ_0 to $\lambda_L \approx 0$ on a log-scale. Note that $\hat{\boldsymbol{\beta}}(\lambda_0) = \mathbf{0}_p$, where λ_0 is given in (3). By default, we use $\lambda_L = \epsilon \lambda_0$, so $\lambda_j = \epsilon^{j/L} \lambda_0 = \epsilon^{1/L} \lambda_{j-1}$ with $\epsilon = 10^{-3}$ which is the default value used by Lasso routine of MATLAB$^{\copyright}$. We consider a dense grid of $(L = 100)$ values of λ.

The conventional approach then uses the information criterion to choose the best solution among these candidate solutions as follows. First, we compute the scale estimate of each solution as:

$$\hat{\sigma}_u^2(\lambda_l) = \frac{1}{n - \|\hat{\boldsymbol{\beta}}(\lambda_l)\|_0} \|\mathbf{y} - \boldsymbol{\Phi}\hat{\boldsymbol{\beta}}(\lambda_l)\|_2^2, \quad l = 0, 1, \ldots, L.$$

Then we compute the GIC optimum (for all values of γ)

$$\ell = \underset{l \in \{0,1,\ldots,L\}}{\arg\min} \{ \text{GIC}_\gamma(\lambda_l) = n \ln \hat{\sigma}_u^2(\lambda_l) + \|\hat{\boldsymbol{\beta}}(\lambda_l)\|_0 \, c_{n,\gamma} \}$$

and choose $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\lambda_\ell)$ as GIC based Lasso solution which gives $\hat{k}^* = \|\hat{\boldsymbol{\beta}}(\lambda_\ell)\|_0$.

III. COMPRESSED BEAMFORMING APPLICATION

A. Single-snapshot Source Localization Problem

Consider a sensor array processing application in which a uniform linear array (ULA) of n sensors is used for estimating the direction-of-arrivals (DoA-s) of the sources with respect to the array axis. The array response (steering vector) of ULA for a source from DoA (in radians) $\theta \in [-\pi/2, \pi/2]$ is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{n}} \left(1, e^{j\pi \sin\theta}, \ldots, e^{j\pi(n-1)\sin\theta} \right)^\top,$$

where we assume half a wavelength inter-element spacing between sensors. We consider the case that $k^* < n$ sources from distinct DoAs $\theta_1, \ldots, \theta_{k^*}$ arrive at a sensor at some time instant t. A single snapshot obtained by ULA can then be modeled as

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{s} + \boldsymbol{\varepsilon} \qquad (8)$$

where $\mathbf{s} \in \mathbb{C}^{k^*}$ contains the source waveforms at time instant t, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{k^*})^\top$ collects the DoAs, and $\mathbf{A} =$

$(\mathbf{a}(\theta_1)\cdots\mathbf{a}(\theta_{k^*})) \in \mathbb{C}^{n\times k^*}$ is the array steering matrix and $\boldsymbol{\varepsilon} \in \mathbb{C}^n$ is the complex random noise.

Consider an angular grid of size p (commonly $p \gg n$) of look directions of interest in terms of the broadside angles,

$$[\theta] = \{\theta_{[j]} \in [-\pi/2, \pi/2) \; : \; \theta_{[1]} < \cdots < \theta_{[p]}\}.$$

Let the jth column of the measurement matrix $\boldsymbol{\Phi}$ in the model (1) be the array response for look direction $\theta_{[j]}$, so $\phi_j = \mathbf{a}(\theta_{[j]})$. Then if the true source DoAs are contained in the angular grid, i.e., $\theta_k \in [\theta]$ for $k = 1, \ldots, k^*$, then the snapshot $\mathbf{y}$ in (8) can be equivalently modeled as in (1) where $\boldsymbol{\beta}$ is exactly k^*-sparse (i.e., $\|\boldsymbol{\beta}\|_0 = k^*$) and non-zero elements of $\boldsymbol{\beta}$ maintain the source waveforms $\mathbf{s}$, i.e., $\boldsymbol{\beta}_{\mathcal{A}^*} = \mathbf{s}$, where $\mathcal{A}^* = \text{supp}(\boldsymbol{\beta})$. Thus, identifying the true DoA-s is equivalent to identifying the non-zero elements of $\boldsymbol{\beta}$. This principle is known as *compressed beamforming (CBF)*. Hence, based on a single snapshot only, c-LARS-GIC can be utilized for detecting the number of sources k^* and the corresponding source parameters, such as DoAs as locations in the grid $[\theta]$ specified by $\mathcal{A}^*$ and the source powers by $|s_k| = \left|[\boldsymbol{\beta}_{\mathcal{A}^*}]_k\right|$ for $k = 1, \ldots, k^*$.

B. Simulation Studies and Results

In CBF, we apply the developed methods for estimating the true source parameters, e.g., number of sources (i.e., sparsity order) k^* and their DoAs and powers. In this paper, we have grid $[\theta]$ for a grid-spacing $\Delta\theta = 2°$ which gives $p = 90$ possible DoAs in $\theta \in [-90°, 90°)$. The source waveforms $\mathbf{s}$ are generated as $s_k = |s_k| \cdot e^{j\text{Arg}(s_k)}$ for $k = 1, \ldots, k^*$, where source powers $|s_k| \in (0,1]$ are fixed but the source phases are randomly generated for each of M Monte-Carlo (MC) trials as $\text{Arg}(s_k) \sim \text{Unif}(0, 2\pi)$. The error terms ε_i are independent and identically distributed and generated from $\mathbb{CN}(0, \sigma^2)$ distribution, where σ^2 is chosen to yield desired signal-to-noise ratio (SNR) in decibels (dB), $\text{SNR(dB)} = 10\log_{10}(\sigma_s^2/\sigma^2)$, where $\sigma_s^2 = \frac{1}{k^*}\{|s_1|^2 + |s_2|^2 + \cdots + |s_{k^*}|^2\}$ denotes the average source power.

The performance measures are the (empirical) *probability of detection*, $\text{PD} = \text{ave}\{\mathsf{I}(k^* = \hat{k}^*)\}$ and the empirical probability of exact recovery, $\text{PER} = \text{ave}\{\mathsf{I}(\mathcal{A}^* = \hat{\mathcal{A}}^*)\}$. We also report the mean squared error, $\text{MSE} = \text{ave}\{\|\mathbf{s} - \hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}^*}\|_2^2\}$ to illustrate the performance in estimation of the source powers. Above the average is over $M = 1000$ MC trials and $\hat{k}^*$, $\hat{\mathcal{A}}^*$ and $\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}^*}$ denote the estimates of k^*, $\mathcal{A}^*$ and $\boldsymbol{\beta}_{\mathcal{A}^*}$ obtained by running the Algorithm 1 for a given MC data set.

For first simulation setup, upper panel in Fig. 1 depicts the full Lasso solution path with found knots for a snapshot generated from a model, where there exists $k^* = 6$ true sources located at $\boldsymbol{\theta} = (-8, -2, 10, 32, 40, 62)°$ with respective powers $(|s_1|, \ldots, |s_6|) = (0.6, 1, 0.9, 1, 0.6, 0.3)$. In this scenario, we have an SNR level of 15 dB and $n = 30$ number of sensors in the ULA. The lower left panel of Fig. 1 depicts the plot of GIC scores $\text{GIC}_0(\lambda_k)$. As can be seen the c-LARS-GIC method chooses a model that has sparsity order $\hat{k}^* = k^* = 6$. The estimated source powers, $|\hat{\boldsymbol{\beta}}_{\hat{\mathcal{A}}^*}|$, at DoAs corresponding to found $\hat{\mathcal{A}}^*$ are also shown in lower right panel of Fig. 1.
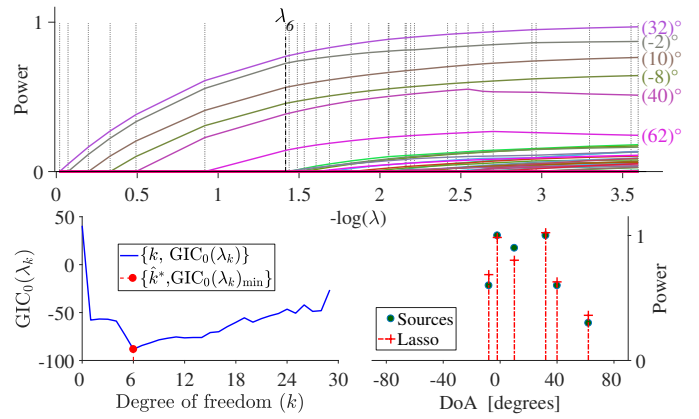

Fig. 1: Upper panel shows the Lasso solution path, i.e., $|[\hat{\boldsymbol{\beta}}(\lambda)]_j|$, with values of the knots λ_k as dotted vertical lines, for $k = 0, 1, \ldots, K = 29$. Lower left panel shows the GIC scores for $\gamma = 0$ as a function of k. The minimum GIC-score is obtained at $k = 6$ giving $\hat{k}^* = 6$ and the magnitudes of the respective $\hat{k}^*$-sparse estimate are shown in lower right panel.
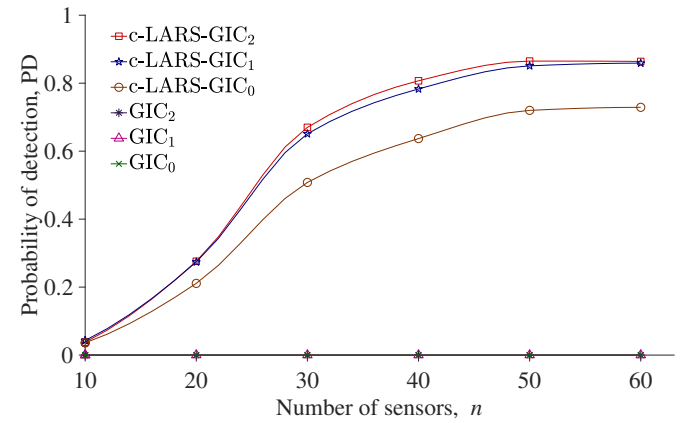

Fig. 2: Probability of detection of sparsity order $k^* = 3$ for varying number of sensors n. The SNR level is 15 dB.

Observe that estimated $\hat{\mathcal{A}}^*$ coincides with the true $\mathcal{A}^*$ and that the source powers are estimated with high accuracy.

In the second simulation we have $k^* = 3$ true sources at $\boldsymbol{\theta} = (-8, 6, 24)°$ with respective powers $(|s_1|, |s_2|, |s_3|) = (0.7, 0.9, 1)$ and the SNR level is 15 dB. When the number of sensors in the ULA is $n = 40$, then the c-LARS-GIC$_\gamma$ had PD equal to 0.640, 0.791, and 0.800 for $\gamma = 0, 1, 2$, respectively. However, conventional grid-based GIC$_\gamma$, as explained in II-D, failed completely yielding miserable PD $= 0$ in all cases. Fig. 2 shows the PD as a function of number of sensors n in the ULA.

Next in the third simulation study, Fig. 3 shows variation in PD according to the number of non-zeros (i.e., sources) for fixed $n = 40$. An SNR level of 20 dB is used in each scenario when k^* ranges from $k^* = 1, \ldots, 10$. Herein, a new source at different (straight or oblique) DoA is introduced randomly as the value of k^* increases. We have sources in the following order $\boldsymbol{\theta} = (30, -14, -22, -32, 16, -2, 56, -30, -8, 58)°$ having corresponding magnitudes $(|s_1|, \ldots, |s_{10}|) =$
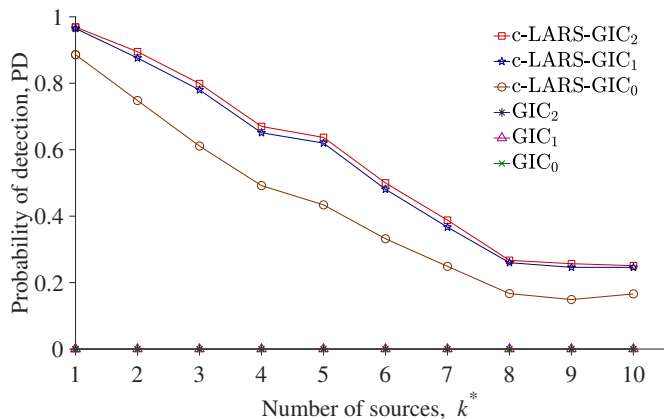
Fig. 3: Probability of detection (PD) for varying number of sources k^*. The SNR level is 20 dB and $n = 40$.
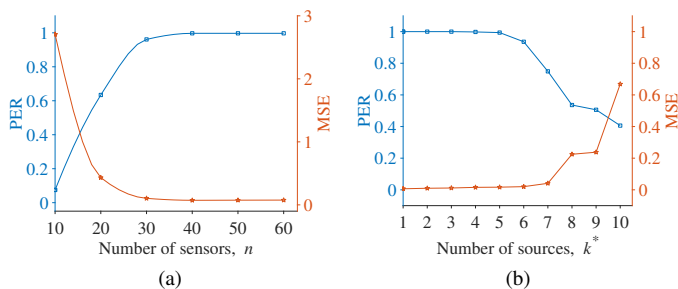

(a) (b)

Fig. 4: Probability of exact recovery (PER) of the true support $\mathcal{A}^*$ and the mean squared error (MSE) for (a) varying number of sensors n and (b) varying number of sources, i.e., sparsity order k^*.

$(0.8, 0.6, 0.4, 0.6, 0.3, 0.2, 0.9, 0.8, 0.4, 0.9)$. It is noticeable that PD decreases as k^* increases. This is expected, since we no longer have $k^* \ll n$ as n is fixed through this simulation. Moreover, the addition of new sources, especially at oblique angles, results in increased mutual coherence which lowers estimation accuracy [8]. These aspects account for the variation in both probabilities PD and PER of c-LARS-GIC for changing n and sparsity order k^*, as visible in both Fig. 3 and Fig. 4 which display the respective changes in PER and MSE for second and third simulation setups, respectively. Note, however, that the conventional grid-based GIC approach fails to detect the true sparsity level in all the cases.

IV. DISCUSSIONS AND CONCLUSIONS

In this paper, we developed the c-LARS based GIC method for finding the sparsity order k^* and corresponding k^*-sparse signal estimate. We compared the c-LARS-GIC method to conventional grid-based GIC approach in single-snapshot source localization based on compressed beamforming.

c-LARS-GIC finds precise values of the penalty parameter where a new variable enters the Lasso path and computes the GIC values for the corresponding active sets. The conventional grid-based GIC approach almost always failed as it heavily depends upon the chosen grid. Simulation studies illustrated that c-LARS-GIC method attains high detection rates even at low SNR level of 15 dB. Moreover, the PER and MSE values follow the same pattern and illustrate that c-LARS-GIC is also able to perform accurate sparse signal reconstruction (SSR).

In conclusion, the results illustrate the potential usage of c-LARS-GIC method in detecting the rank of the signal subspace, i.e., accurate sparsity (model) order selection. Furthermore, its ability to find the true support and magnitudes estimation in SSR task makes it favorable option. Finally, our software package, containing c-LARS-GIC code along with algorithms of [8], is freely available at [9].

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations*. CRC Press, 2015.

[3] F. Römer, A. Lavrenko, G. Del Galdo, T. Hotz, O. Arikan, and R. S. Thomä, "Sparsity order estimation for single snapshot compressed sensing," in *48th Asilomar Conference on Signals, Systems and Computers, 2014*. IEEE, 2014, pp. 1220–1224.

[4] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, "On the relation between sparse reconstruction and parameter estimation with model order selection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 560–570, 2010.

[5] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[7] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression (with discussion)," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[8] M. N. Tabassum and E. Ollila, "Sequential adaptive elastic net approach for single-snapshot source localization," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, 2018.

[9] ——, "Software package — doi: 10.5281/zenodo.1238450," 2018. [Online]. Available: www.github.com/mntabassm/SAEN-LARS (Last viewed 15 June 2018).

[10] ——, "Pathwise least angle regression and a significance test for the elastic net," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, Aug 2017, pp. 1309–1313.

[11] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[12] V. Koivunen and E. Ollila, *Academic Press Library in Signal Processing, Volume 3: Array and Statistical Signal Processing*. Elsevier, 2014, ch. Model order selection, pp. 9–26.

[13] R. Lockhart, J. Taylor, R. J. Tibshirani, and R. Tibshirani, "A significance test for the lasso," *Ann. Statist.*, vol. 42, no. 2, pp. 413–468, 04 2014.

[14] R. J. Tibshirani, J. Taylor, R. Lockhart, and R. Tibshirani, "Exact post-selection inference for sequential regression procedures," *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 600–620, 2016.

[15] R. Nishii, "Asymptotic properties of criteria for selection of variables in multiple regression," *Ann. Statist.*, vol. 12, no. 2, pp. 758–765, 06 1984.

[16] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[17] H. Wang, B. Li, and C. Leng, "Shrinkage tuning parameter selection with a diverging number of parameters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 3, pp. 671–683, 2009.

[18] Y. Fan and C. Y. Tang, "Tuning parameter selection in high dimensional penalized likelihood," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 75, no. 3, pp. 531–552, 2013.