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

I. INTRODUCTION

We assume that a measurement vector \( y \in \mathbb{C}^n \) is generated via linear model
\[
\begin{align*}
y = \Phi \beta + \varepsilon,
\end{align*}
\]
where \( \Phi = (\phi_1, \ldots, \phi_p) \in \mathbb{C}^{n \times p} \) is the known basis matrix (or matrix of predictors), \( \beta \in \mathbb{C}^p \) is the unknown signal vector (or regression coefficient vector) and \( \varepsilon \in \mathbb{C}^n \) is the (unobserved) random noise vector. We assume that the model is underdetermined, i.e., \( p > n \), and the signal vector \( \beta \) is sparse (i.e., having only a few non-zero elements) with sparsity order \( k^* = \| \beta \|_0 \ll n \), where \( \| \cdot \|_0 \) denotes the \( \ell_0 \) (pseudo)norm, \( \| \beta \|_0 = |\{ i : \beta_i \neq 0 \}| \), i.e., \( \| \beta \|_0 \) is equal to number of non-zero elements of \( \beta \). Such a model arises in compressed sensing [1] and in sparse linear regression [2].

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 \( \beta \).

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 \( \lambda \) commonly chosen by cross-validation to compute an estimate \( \hat{\beta} \) of sparse vector \( \beta \). An estimate of the sparsity order \( k^* \) is then obtained as \( \hat{k}^* = \| \hat{\beta} \|_0 \).

The Lasso estimator of \( \beta \) solves the following penalized (regularized) residual sum of squares (RSS) optimization problem,
\[
\hat{\beta}(\lambda) = \arg \min_{\beta \in \mathbb{C}^p} \frac{1}{2} \| y - \Phi \beta \|_2^2 + \lambda \| \beta \|_1
\]

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 \( \lambda \) varies. The LARS method finds the sequence of values \( \lambda_k \), 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 \( \beta(\lambda_k) \) 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 ℓ2 and ℓ1 norms of the vector a ∈ ℂp are defined as ∥a∥2 = √a∗a and ∥a∥1 = ∑j=1p |a|j, where |a| = √a∗a = √aR+aI denotes the modulus of a complex number a = aR + jaI and (·)H = [(·)∗]T denotes the Hermitian (complex conjugate) transpose. The support of a is the index set of its non-zero elements, i.e., A = supp(a) = {j ∈ {1, . . . , p} : a_j ̸= 0}. The ℓ0-(pseudo)norm of a is defined as ∥a∥0 = |supp(a)|, which is equal to the total number of non-zero elements in it. If β ∈ ℂp, Φ ∈ ℂn×p and A is an index set A ⊂ {1, . . . , n} of cardinality |A| = k, then βA ∈ ℂk (resp. ΦA) denotes the sub-vector (resp. n × k sub-matrix) restricted to components of β (resp. columns of Φ) indexed by the set A. The Φ+ is a Moore-Penrose pseudoinverse of Φ. We use l(·) to denote the indicator function and ave = 1M ∑M=m=1(·) for M numbers.

II. THE c-LARS-GIC METHOD

A. The Lasso Knots and Active Sets

Recall that the signal vector β ∈ ℂp is sparse, so only few elements are non-zero. The support A* of β is the index set of its non-zero elements, i.e., A* = supp(β) and the cardinality of this set, k* = ∥β∥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 A*. Naturally, if A* would be known, then one can simply compute the LS solution when redefining y on ΦA*, where ΦA* denotes the n × k* (sub)matrix of Φ restricted to the columns of the active set A*.

Recall that Lasso estimate ˆβ(λ) ∈ ℂp is sparse having at most n non-zero elements. The sparsity level of ˆβ(λ) depends on the penalty parameter λ. We assume that the columns of Φ are centered, i.e., ∥φj∥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 φj-s; see [2] for more elaborate discussion of this feature of the Lasso.

Let λ0 denotes the smallest value of λ such that all coefficients are zero, i.e., β(λ0) = 0p. It is easy to see that [2]

$$\lambda_0 = \max_{j \in \{1, \ldots, p\}} |\phi_j y|. \quad (3)$$

Let A(λ) = supp(β(λ)) denotes the active set at the regularization parameter value λ < λ0. The knots λ0 > λ1 > λ2 > ⋯ > λK 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 λk is denoted by A_k = A(λ_k) = supp(β(λ_k)). The active set A_1 thus contains a single index as A_1 = {j_1}, where j_1 is predictor that becomes active first, i.e.,

$$j_1 = \arg \max_{j \in \{1, \ldots, p\}} |\phi_j y|. \quad \text{By definition of the knots, one has that}$$

$$A_k = \{\lambda_k \leq \lambda < \lambda_{k+1} \quad \forall \lambda_{k+1}, \forall k = 0, 1, \ldots, K$$

and A_0 = supp(β(λ)) = {0} for all λ ≥ λ_0.

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

$$H_A : \beta_A = 0, \quad |A| = k \Rightarrow |A^c| = p - k \quad (4)$$

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

Finding the Lasso knots λ_0, λ_1, . . . , λ_K and the corresponding active sets A_k for k = 0, 1, . . . , K, via the c-LARS-WLasso algorithm of [8], [9] with unit weights allows us to form a nested set of hypotheses,

$$H_0 \subset H_1 \subset \cdots \subset H_K$$

$$H_k : \beta_{A_k} = 0, \quad k = 0, 1, \ldots, K. \quad (5)$$

This reduces the number of tested hypotheses from n^n to (K + 1) ≤ n ≤ p.
Algorithm 1: The c-LARS-GIC method.

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

1 Compute the knots \( \lambda_k \) and the corresponding active sets \( A_k \) 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:
\[
\hat{k}^* = \arg\min_{k \in \{0, 1, \ldots, K\}} \{ \text{GIC}_n(\lambda_k) = n \ln \hat{\sigma}_n^2(A_k) + k c_{n, \gamma} \},
\]
where \( \hat{\sigma}_n^2(A_k) \) is computed using (7). Thus, an estimate of the active set is \( \hat{A}^* = A_{\hat{k}^*} \).

3 Set \( \beta_{\hat{A}^*}^{} = (\Phi_{\hat{A}^*}^\dagger)^{\dagger} y \) to get \( k^* \)-sparse estimate \( \hat{\beta} \in \mathbb{C}^p \).

where the sequence \( \{\lambda_l\} \) is monotonically decreasing from \( \lambda_0 \) to \( \lambda_L \approx 0 \) on a log-scale. Note that \( \hat{\beta}(\lambda_0) = 0_p \), where \( \lambda_0 \) is given in (3). By default, we use \( \lambda_L = \epsilon_0 \), so \( \lambda_l = \epsilon_l/L \lambda_0 = \epsilon_l/L \lambda_{l-1} \) with \( \epsilon = 10^{-3} \) which is the default value used by Lasso routine of MATLAB\textsuperscript{®}. We consider a dense grid of \( (L = 100) \) values of \( \lambda \).

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:
\[
\hat{\sigma}_n^2(\lambda_l) = \frac{1}{n - \|\hat{\beta}(\lambda_l)\|_0^2} \| y - \Phi \hat{\beta}(\lambda_l) \|_2^2, \quad l = 0, 1, \ldots, L.
\]
Then we compute the GIC optimum (for all values of \( \gamma \))
\[
\ell = \arg\min_{l \in \{0, 1, \ldots, L\}} \{ \text{GIC}_n(\lambda_l) = n \ln \hat{\sigma}_n^2(\lambda_l) + \|\hat{\beta}(\lambda_l)\|_0 c_{n, \gamma} \}
\]
and choose \( \hat{\beta} = \hat{\beta}(\lambda_{\ell}) \) as GIC based Lasso solution which gives \( \hat{k}^* = \|\hat{\beta}(\lambda_{\ell})\|_0^2 \).

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 \( \theta \) in radians \( \theta \in [-\pi/2, \pi/2] \) is given by
\[
a(\theta) = \frac{1}{\sqrt{n}} \left( 1, e^{j\pi \sin \theta}, \ldots, e^{j\pi (n-1) \sin \theta} \right)^T,
\]
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
\[
y = A(\theta)s + \epsilon
\]
where \( s \in \mathbb{C}^{k^*} \) contains the source waveforms at time instant \( t \), \( \theta = (\theta_1, \ldots, \theta_{k^*})^T \) collects the DoAs, and \( A = \)
\((a(\theta_1) \cdots a(\theta_k)) \in \mathbb{C}^{n \times k} \) is the array steering matrix and \(\varepsilon \in \mathbb{C}^n\) 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,

\[ [\theta] = \{\theta_{[j]} \mid \theta_{[j]} \in [-\pi/2, \pi/2] : \theta_{[j]} < \cdots < \theta_{[\|]} \}. \]

Let the \(j^{th}\) column of the measurement matrix \(\Phi\) in the model (1) be the array response for look direction \(\theta_{[j]}\), so \(\phi_j = a(\theta_{[j]})\). 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 \(y\) in (8) can be equivalently modeled as in (1) where \(\beta\) is exactly \(k^*\)-sparse (i.e., \(|\beta|_0 = k^*\)) and non-zero elements of \(\beta\) maintain the source waveforms \(s\), i.e., \(\beta_{A^*} = s\), where \(A^* = \text{supp}(\beta)\). Thus, identifying the true DoA-s is equivalent to identifying the non-zero elements of \(\beta\). 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 \(A^*\) and the source powers by \(|s_k| = |\beta_{A^*}|_{k}\) 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 \(\Delta \theta = \alpha\) which gives \(p = 90\) possible DoAs in \([\theta] \in [-90^\circ, 90^\circ]\). The source waveforms \(s\) are generated as \(s_k = |s_k| \cdot e^{j \arg(s_k)}\) 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 \(\arg(s_k) \sim \text{Unif}(0, 2\pi)\). The error terms \(\varepsilon_i\) are independent and identically distributed and generated from \(\mathcal{CN}(0, \sigma^2)\) distribution, where \(\sigma^2\) is chosen to yield desired signal-to-noise ratio (SNR) in decibels (dB), \(\text{SNR}(\text{dB}) = 10 \log_{10}(\sigma^2 / \sigma^2)\), where \(\sigma^2 = 1/F \left( |s_1|^2 + |s_2|^2 + \cdots + |s_{k^*}|^2 \right)\) denotes the average source power.

The performance measures are the (empirical) probability of detection, \(PD = \text{ave}(I[k^* = k^*])\) and the empirical probability of exact recovery, \(\text{PER} = \text{ave}(I[A^* = A^*])\). We also report the mean squared error, \(\text{MSE} = \text{ave}(\|s - \hat{A} \cdot s\|^2)\) to illustrate the mean squared error in estimation of the source powers. Above the average is over \(M = 1000\) MC trials and \(k^*, A^*\) and \(\beta_{A^*}\) denote the estimates of \(k^*, A^*\) and \(\beta_{A^*}\) 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 \(\theta = (\pm 8, \pm 2, 10, 32, 40, 62)\) with respective powers \((|s_1|, \ldots, |s_{10}|) = (0.6, 1, 0.9, 1, 0.6, 0.3)\). 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 \(\text{GIC}_0(\lambda)\). As can be seen the c-LARS-GIC method chooses a model that has sparsity order \(k^* = k^* = 6\).

The estimated source powers, \(|\beta_{A^*}|\), at DoAs corresponding to found \(A^*\) are also shown in lower right panel of Fig. 1.

![Fig. 1: Upper panel shows the Lasso solution path, i.e., \(|\beta(\lambda)|_j\), with values of the knots \(\lambda\) 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 \(k^* = 6\) and the magnitudes of the respective \(k^*\)-sparse estimate are shown in lower right panel.](image1.png)

![Fig. 2: Probability of detection of sparsity order \(k^* = 3\) for varying number of sensors \(n\). The SNR level is 15 dB.](image2.png)
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].

ACKNOWLEDGMENT

The research was partially supported by the Academy of Finland grant no. 298118 which is gratefully acknowledged.

REFERENCES